

Structural Studies of Icosahedral Viruses

Hanwen Zhang

A thesis submitted in partial fulfilment of the requirements for the degree
of

Doctor of Philosophy



St-Hugh's College
Nuffield Department of Medicine
Division of Structural Biology
University of Oxford

Michaelmas 2017

Abstract

Described as “organisms at the edge of life”, viruses are the most abundant type of biological entity that can infect all types of life forms and are found in almost every ecosystem on Earth. Although virus capsids can be found with a wide range of shapes, around half of the virus families have an icosahedral arrangement. This thesis mainly focuses on structural studies of proteins from two types of viruses: the VP5 protein of Bluetongue Virus (BTV) serotype 15, and the capsid protein (CP) of Maize Streak Virus (MSV) (the Nigerian strain).

Bluetongue (BT) disease in ruminates is caused by BTV and is of important agricultural and economic impact. As one of the two proteins that comprise the outer capsid of the virus, VP5 has been shown to play an important role in membrane penetration. I describe structures of BTV VP5, at both high and low pH, both solved by X-ray crystallography to atomic resolution. These provide insights into the membrane penetration process for a non-enveloped virus at the molecular level. An atomic level mechanism of pH dependant conformation changes was revealed, including a re-arrangement of 3 bent helices into one long straight helix protruding out of the molecule, and a transformation from a β -sheet to an α -helix. The significant change due to the movement of the helix of the so-called unfurling domain could be caused by a conformation change of H386 pushing the conserved E316 residue in the so-called beam helix, α 12, which contacts the anchoring domain. Our low pH structure fits well into the low pH cryo-EM image, clarifying the observed protruding density.

Geminiviruses cause diseases in various plants, leading to great economic loss. MSV is the type member of genus *Mastrevirus* of the family *Geminiviridae*. The MSV capsid is

composed of a single protein, termed, CP. Infectious virus particle capsids are generally constructed from two incomplete T=1 icosahedra, joined to form a so-called geminated viral particle of dimensions 220 x 380 Å with a fivefold symmetry axis coincident with the long axis, so that the CPs are arranged with quasi-icosahedral symmetry. The MSV VLP (single icosahedral particles) and the geminated MSV virion structures were solved to resolutions of 9.5 Å and 7.4 Å, respectively, by cryo-EM. A localized reconstruction method was used in the processing procedure in the geminated MSV virion structure by treating each half of the geminated particle (one incomplete icosahedral half) as a new single particle. These structures revealed structural information about the virus and help our understanding of the assembly of geminiviruses. The MSV geminated virus particles are characterized by a twinned icosahedral capsid with distorted peripentonal pentamers and equatorial pentamers observed by us, and a 12° offset between the two halves of the particle.

Acknowledgements

First and foremost, I would like to thank my supervisor, Professor Dave Stuart for his guidance and support in the last four years. His passion for research has been exceptionally inspiring, which has made the project more enjoyable. I would also like to thank Dr Geoff Sutton, who has provided day-to-day supervision to me. In particular, Geoff has been demonstrating and explaining a lot to me in the lab, from the tiny experimental details to fundamental principles. I greatly appreciate his patience and guidance.

I also wish to express sincere thanks to other members in the lab, particularly Jingshan Ren, Yuguang Zhao, Kamel El Omari, Karl Harlos, Abhay Kotacha, and Tom Walter, for their generosity in supporting me and helping out. Their knowledge, guidance and advice in many aspects has made this thesis possible. Besides, I would like to thank Margaret and Weixian, for their support and dedication to make sure the day-to-day running of the lab.

It has been a particularly memorable experience at STRUBI in the last four years, together with a group of open-minded and warm-hearted people. I am very thankful to all my friends here, Youxin, Holly, Sai, especially Tao for accompanying with me to share the joys and gossips, and overcome frustrations.

Finally, I wish to thank my parents Chunlei Zhang, Liping Zhang, my boyfriend Xiaoguang Li for being the pillars of my life, they have been keeping me motivated throughout the four years of doctoral study. Also I want to thank my dear bros Ying, Lofty and Johnsky for their encouragement and support during the last few months of my PhD journey.

"The stars may look like ants, but from the deep of your heart, you know they are different."

—H. Zhang

Contents

Abstract	I
Acknowledgements	III
Contents.....	1
List of Figures	6
List of Tables	12
Abbreviations.....	13
Chapter 1. Introduction.....	15
1.1. Structural studies of viruses	16
1.1.1. Very brief introduction to viruses	16
1.1.2. Structural study of viruses using X-ray crystallography.....	19
1.1.3. 1.1.3. Application of virus structures	20
1.2. Structural Studies of Bluetongue Virus VP5	20
1.2.1. Brief Introduction of Bluetongue Disease	20
1.2.2. Serotypes of BTV	22
1.2.3. Transmission and Controlling of BTV	24
1.3. Overview of Structural Studies on Bluetongue Virus and Protein Functions in the Virus Life Cycle	25
1.3.1. Proteins of BTV	25
1.3.2. The structural proteins of BTV and their functions	26
1.3.3. The non-structural proteins of BTV and their functions	32

1.3.4.	Non-enveloped viruses (especially BTV) cell entry.....	36
1.3.5.	Structure features of BTV related to cell entry	40
1.4.	Structural and Functional Studies of Maize Streak Virus	42
1.4.1.	Introduction of Geminiviruses.....	42
1.4.2.	The geminiviruses infection of plants	45
1.4.3.	Genimivirus interaction with the insect hosts.....	47
1.4.4.	Introduction of the capsid protein (CP) of geminiviruses	48
1.4.5.	Introduction of ther proteins in geminiviruses	51
1.4.6.	Introduction to Maize Streak Viruses.....	56
1.4.7.	Introduction of the CP of MSV and the structure of the viral capsid	59
Chapter 2.	Materials and Methods.....	63
2.1.	Constructs Design	64
2.1.1.	Protein sequence alignment and phylogenetic analysis of 26 serotypes of BTV	64
2.1.2.	Constructs design	64
2.2.	High-throughput cloning	65
2.2.1.	Primer design	66
2.2.2.	DNA amplification by polymerase chain reaction (PCR)	66
2.2.3.	High-throughput cloning with InFusion™ system	68
2.2.4.	Transformation of plasmid/ InFusion™ mixture into competent cells and PCR verification.....	68
2.3.	Protein expression in <i>E. coli</i>	69
2.4.	SDS-PAGE and western blot analysis	69
2.4.1.	SDS-PAGE analysis	69
2.4.2.	Western blot analysis	70
2.5.	MSV Capsid Protein Purification from <i>E. coli</i>	71

2.6.	Protein Expression in insect cells (<i>Sf9</i> cells)	72
2.6.1.	<i>Sf9</i> Cell Culture	72
2.6.2.	Generation of recombinant baculoviruses.....	72
2.6.3.	Preparation of Stock Baculoviruses.....	72
2.6.4.	BTV15 VP5 Protein Expression	73
2.6.5.	Selenomethionine-incorporated protein expression	73
2.7.	Protein purification from <i>Sf9</i> cells	73
2.7.1.	Cell harvest and homogenizing	73
2.7.2.	Primary purification by binding to Cobalt beads.....	74
2.8.	Purification of maize streak virus from plants	74
2.9.	Crystallization	75
2.9.1.	Crystallization Screens.....	75
2.9.2.	Crystal Optimizations	77
2.10.	X-ray crystallography data collection and processing	77
2.11.	Cryo-protection and crystal mounting	77
2.11.1.	Data collection	78
2.11.2.	Data processing	78
2.12.	Crystal Structure determination	79
2.12.1.	Phasing by molecular replacement	79
2.12.2.	Refinement.....	80
2.12.3.	Structure validation.....	81
2.13.	Transmission electron microscopy data collection.....	81
2.13.1.	Basic principles of negative stain electron microscopy (EM)	81
2.13.2.	Sample preparation for negative stain EM.....	83
2.13.3.	Sample checking on Tecnai T12 Electron Microscopy.....	83
2.13.4.	Graphene oxide grid preparation.....	84

2.13.5.	Sample preparation for cryo-EM with Vitrobot	84
2.14.	Data collection, processing and refinement for maize streak viruses for cryo-EM	86
2.14.1.	Cryo-EM data collection	86
2.14.2.	Basic principles of cryo-EM data processing in Relion	86
2.14.3.	Localized reconstruction on two halves of geminated MSV viruses	89
2.15.	Multi-angle light scattering analysis	90
Chapter 3.	Results of BTV VP5 structures	91
3.1.	Expression constructs screen for bluetongue virus VP5	92
3.2.	Protein expression and purification.....	96
3.2.1.	Baculovirus Expression system.....	96
3.2.2.	Construct design and cloning	96
3.2.3.	Constructs expression and solubility test.....	101
3.2.4.	Large scale protein expression and protein purification of BTV15 VP5	104
3.2.5.	Protein characterizations by MultiAngle Light Scattering (MALS).....	106
3.3.	Protein Crystallization	108
3.4.	Data collection and processing.....	109
3.5.	The structure of BTV15 VP5 at high pH.....	114
3.5.1.	The overall structure	117
3.5.2.	Comparison of crystal structure and cryo-EM structures	121
3.5.3.	VP5 is a trimer in structure.....	123
3.5.4.	BTV VP5 shows a conserved surface for the subunit interfaces in the trimer, but more variable outer surface.....	125
3.6.	Structure of VP5 at low pH.....	127
3.6.1.	Model building for low pH VP5 structure	127
3.6.2.	Overall structure of low pH VP5 structure	133

3.7.	Locating the pH sensor for conformation change trigger	136
3.8.	Fitting the low pH crystal structures into cryo-EM map.	141
3.9.	Discussion	143
Chapter 4.	Structural Studies of MSV	150
4.1.	Structural Studies of MSV VLPs	151
4.1.1.	Construct design and cloning	151
4.1.2.	Protein expression and purification of MSV-CP	151
4.1.3.	Cryo-EM structure of MSV-VLP	153
4.1.4.	Alignment of MSV and STNV and Fitting of STNV CP into MSV VLP.....	158
4.1.5.	Resmap analysis of MSV-VLP EM structure.....	160
4.2.	Structural Studies of MSV geminated virus particles	161
4.2.1.	MSV Virus purification and preliminary EM sample analysis	161
4.2.2.	Cryo-EM structure of MSV-gemini via single particle reconstruction	164
4.2.3.	Connection between the two halves of the MSV particles	166
4.2.4.	Localized reconstruction of MSV	169
4.2.5.	Rotation offset of the two halves of MSV particles.....	170
4.2.6.	Resmap analysis of MSV localize reconstructed EM structure.....	172
4.2.7.	Fitting STNV CP into localize reconstructed MSV EM density	173
4.3.	Discussion	176
Chapter 5.	Conclusions and Future Work	179
5.1.	Overview	180
5.2.	Conclusions and Future work on BTV VP5 structures.....	181
5.3.	Conclusion and Future Work of the MSV CP structures	183
References		185

List of Figures

Figure 1.1 Some representative viruses shown in the same scale.....	17
Figure 1.2 Symptoms of sheep with bluetongue.....	21
Figure 1.3 Cryo-EM density map of the BTV virion.....	27
Figure 1.4 Structure of VP5 at low pH.	28
Figure 1.5 The core particle of BTV-1.	29
Figure 1.6 Cartoon representation of the X-ray structure of the VP4 molecule.	31
Figure 1.7 Ribbon representation of the NS2N homodimeric configurations and overview of the crystal packing.	34
Figure 1.8 Mechanisms that initiate the membrane penetration process. (Tsai, 2007).	39
Figure 1.9 Phylogenetic tree of the full genome sequences of representative geminiviruses (Varsani et al., 2014).	44
Figure 1.10 Genome organisation of isolates in various geminivirus lineages.	46
Figure 1.11 Structural modeling of the geminiviral capsid proteins (CPs).	50
Figure 1.12 Schematic summary of viral DNA replication forms. (Preiss & Jeske, 2003).	52
Figure 1.13 Models and image reconstruction of MSV-N.	61
Figure 1.14 Views of Geminate and Single MSV Capsids Down Symmetry Axes..	62
Figure 2.1 Helical wheel representations of (A) aa 1 to 20 and (B) aa 21 to 41 of BTV15 VP5.	65

Figure 2.2 Schematic graph of a transmission electron microscope.....	82
Figure 3.1 Sequence alignments of 26 serotypes of BTV.	94
Figure 3.2 Phylogenetic grouping bluetongue virus VP5.....	95
Figure 3.3 The pOPINE baculovirus transfer vector map.	98
Figure 3.4 Expression and solubility test of BTV serotype 3, 4, 8, 19, 21 and 26 VP5 proteins by SDS-PAGE analysis of (A1-3) whole cells and (B1-3) supernatant.	102
Figure 3.5 Expression and solubility tests of BTV10 and BTV15 VP5 proteins.	103
Figure 3.6 BTV15-VP5- Δ 43-Nhis protein expression with Sf9 cells and purification with a Cobalt beads affinity column and an S200 size exclusion column.....	105
Figure 3.7 Se-Met labelled BTV15-VP5- Δ 43-Nhis protein expression with Sf9 cells and purification with a S200 size exclusion column.	106
Figure 3.8 MALS of the protein from construct BTV15-VP5- Δ 43-Nhis.....	107
Figure 3.9 BTV15 VP5 protein crystals and collected data images.	110
Figure 3.10 X-ray fluorescence scan plot from the Se-Met labelled VP5 crystal. ..	111
Figure 3.11 Crystal morphology and diffraction pattern of the crystal formed from the low pH condition.....	112
Figure 3.12 Anomalous density map for Se-Met in BTV15 VP5.....	115
Figure 3.13 Detailed anomalous density map with labelled Se-Met numbers in BTV15 VP5.	116
Figure 3.14 BTV-15 VP5 amino acid sequence annotation.	118
Figure 3.15 Cartoon representation of BTV-15 high pH VP5 structure.....	119

Figure 3.16 Cartoon and surface view of BTV-15 high pH VP5 structure.	120
Figure 3.17 Alignment of VP5 crystal structure and cryo-EM structure both at high pH.....	122
Figure 3.18 Plot of distance between corresponding C α s of the crystal structure and cryo-EM structure (PDB code: 3J9E).	123
Figure 3.19 BTV VP5 crystal component in a crystallographic unit cell and trimer representation.	124
Figure 3.20 Surface conservation among different serotypes of BTV VP5 proteins.	126
Figure 3.21 The Matthews probability plot for low pH crystal unit cell with all amino acids from the construct.	128
Figure 3.22 Electron density map from low pH crystal data when placed VP5 anchoring domain for phasing.	129
Figure 3.23 The BTV-15 VP5 low pH crystal packing and electron density map. .	130
Figure 3.24 Electron density map (2Fo-Fc) shows an area of low pH VP5 structure during the model building in Coot.	131
Figure 3.25 The BTV-15 VP5 low pH crystal structure refinement process as indicated by cartoon and electron density maps.	131
Figure 3.26 BTV-15 VP5 amino acid sequence and low pH secondary structure annotation.....	132
Figure 3.27 BTV15 VP5 low pH structures presented as monomer and trimers. ...	134
Figure 3.28 Alignment of BTV15 low pH VP5 structures with corresponding high pH structure at two different angle.	135

Figure 3.29 Plot of distance between corresponding C α s of the low pH and high pH structures.	136
Figure 3.30 The close up view of the histidine cluster in VP5 in the area of interface between unfurling domain and anchoring domain.....	137
Figure 3.31 The potential pH sensor Histidine residues H386 and H387 in VP5 structures at different pHs.	139
Figure 3.32 The proposed mechanism of low pH triggered VP5 unfurling domain disorder or detachment.....	140
Figure 3.33 Low pH BTV15 VP5 crystal structure fitted into low pH cryo-EM structures.	142
Figure 3.34 Proposed cartoon model of conformational changes of BTV VP5.	146
Figure 3.35 A cartoon presentation of BTV15-VP5 stem helices at low pH (A) and HIV-1 gp41 heptad repeat 1 (B) and their alignment (C).	149
Figure 4.1 The spectrometry result and the SDS-PAGE test for the gel filtration of MSV CP.....	152
Figure 4.2 MSV-VLP checked with negative stained EM.	153
Figure 4.3 Central slices through the 2D classes of the MSV-VLP particles.	155
Figure 4.4 2D projections of the 20 3D classes of the MSV-VLP particles.	156
Figure 4.5 Final structure of MSV-VLP with a final resolution determined to 9.5 Å.	157
Figure 4.6 Sequence alignment of MSV CP and STNV CP.	159

Figure 4.7 Fitting of one pentamer of the STNV CP into the EM density of a CP pentamer of MSV VLP.	160
Figure 4.8 ResMap analysis of MSV_VLP to colour-code structure based on local resolution. The colours that represents for resolution ranges are: Blue for higher than 4 Å, cyan for 5 Å, green for 6 Å, yellow for 7 Å, orange for 8 Å, red for 9 Å and grey for lower than 9 Å.	161
Figure 4.9 The post-infection maize plants, with streaky leaves and unhealthy growth (A and B).....	162
Figure 4.10 MSV geminated virus particle sample checked with negative stained and cryo-EM.	163
Figure 4.11 Central slices through the 2D classes of the MSV particles.	165
Figure 4.12 2D projections of the four 3D classes of the MSV particles.	165
Figure 4.13 The final resolution of the MSV treated as single particles was determined to 8.3 Å.....	166
Figure 4.14 A gap between the two heads of the MSV particle is shown with density of bridges linking the halves of the virus.....	168
Figure 4.15 Localize reconstructed structure of MSV determined to resolution of 7.4 Å,.....	170
Figure 4.16 Method to illustrate by what degree one half of the particle is offset from the other half.	171
Figure 4.17 ResMap analysis of localize reconstructed structure of half the geminated particle of MSV virus.....	172

Figure 4.18 The fitting of the atomic monomer model from STNV (ribbon) into the EM density (mesh) of MSV virus to generate three types of pentamers shown from both side view and top view of three different capsomers.....	174
Figure 4.19 The fitting of the atomic pentamer models from STNV into the EM density of MSV virus shown from both side view of three different capsomers.	175
Figure 4.20 Inside density of MSV geminated particle (A) and VLP (B) shown at a contour level of 0.0087.	178

List of Tables

Table 2.1 PCR reactions.	67
Table 2.2. Thermal cyclers procedure for amplifying template DNA	67
Table 2.3 Standard crystallization blocks used in this study.	76
Table 3.1 Detail cloning information about BTV VP5 expression constructs.	99
Table 3.2 Data collection and refinement statistics for VP5 crystals at two different pHs.	113
Table 4.1 Correlation coefficient (CC) value of the three types of pentamers generated from STNV CP monomers fitted onto MSV virus localize reconstructed structure using Chimera.	173

Abbreviations

ACMV	African cassava mosaic virus
BCTV	Beet curly top virus
BSCTV	Beet severe curly top virus
BTV-15	Bluetongue Virus serotype 15
CaLCV	Cabbage Leaf Curl Virus
CC	Correlation Coefficient
cccDNA	covalently closed circular DNA
CP	Coat Protein
cryo-EM	cryo-Electron Microscopy
CSR	Complementary-Strand Replication
CTF	Contrast Transfer Function
EM	Electron Microscopy
FBS	Fetal Bovine Serum
hdsDNA	heterogeneous-length dsDNA
IPTG	Isopropyl β -D-1-thiogalactopyranoside
kDa	Kilodalton
LIC	Ligation-Independent Cloning
LIR	Long Intergenic Region
MALS	Multi-Angle Light Scattering
MLV	Modified Live Virus
mM	Millimolar
MoI	Multiplicity of Infection
MP	Movement Protein
MR	Molecular Replacement
MSD	Maize Streak Disease
MSV	Maize Streak Virus
NCS	Non Crystallographic Symmetry
NES	Nuclear Export signal

NS1/2/3/4	None Structural Protein 1/2/3/4
NSP	Nuclear Shuttle Protein
ocDNA	open circular DNA
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PDB	Protein Database Bank
RCR	Rolling-Circle Replication
RDR	Recombination-Dependent Replication
REn	Replication Enhancer Protein
Rep	Replication-associated Protein
SEC	Size Exclusion Chromatography
SeMet	Selenomethionine
<i>Sf</i>	<i>Spodoptera frugiperda</i>
SIR	Short Intergenic Region
siRNAs	small interfering RNAs
SqLCV	Squash Leaf Curl Virus
ssRNA	single-stranded RNA
STNV	Satellite Tobacco Necrosis Virus
TEM	Transmission Electron Microscopy
ToLCJV	Tomato leaf curl Java Virus
ToLCNDV	Tomato leaf curl New Delhi Virus
TrAP	Transcriptional activator protein
TYLCV	Tomato Yellow Leaf Curl Geminivirus
VLP	Virus Like Particles
VP5	Virus Protein 5

Chapter 1.

Introduction

1.1. Structural studies of viruses

1.1.1. Very brief introduction to viruses

Viruses are small structures without enough metabolic pathways and mechanisms to be able to survive on their own and therefore they not only use their own components but also use the cellular machinery in a host cell to replicate their DNA or RNA genome and their protein components in order to allow them to multiply within cells and propagate between cells and organisms (Mateu, 2013b). They are considered as a life form by some because they can reproduce, evolve through natural selection and carry genetic material, while viruses also have been described as “organisms at the edge of life” (Rybicki, 1990) and as replicators (Koonin & Starokadomskyy, 2016) due to their lack of key cell characteristics. There are an estimated 10^{31} viruses on Earth, most of which are phages that infect bacteria (Breitbart & Rohwer, 2005). Viruses can infect all types of life forms, from animals and plants to microorganisms, including bacteria and archaea (Koonin, Senkevich, & Dolja, 2006) and are found in almost every ecosystem on earth and are the most abundant type of biological entity (Delwart, 2007). In 1892, Dmitri Ivanovsky described a non-bacterial pathogen infecting tobacco plants, and this was followed by Martinus Beijerinck discovering in 1898 of the tobacco mosaic virus (Bos, 1999), whilst before the end of the 19th Century the first animal virus, foot-and-mouth disease virus was discovered (Rott & Siddell, 2017). There are millions of types of viruses described among about 5000 detailed species (Breitbart & Rohwer, 2005). In the history of evolutionary life, the origin of viruses is still not clear, there are theories on evolution from plasmids, or from bacteria. The viruses increased the genetic diversity in evolution by means of horizontal gene transfer (Canchaya, Fournous, Chibani-Chennoufi, Dillmann, & Brüssow, 2003).

However there are rather few structural lineages of viruses and the structure of the virus capsid appears to be strongly conserved (Abrescia, Bamford, Grimes, & Stuart, 2012). A famous structural biologist M. Rossmann once wrote, “Modern biology was founded on the study of bacterial viruses and information on virus structures has been growing exponentially since about 1900, with the help from development of experimental and computational techniques, with the growth of molecular biology and virology” (Rossmann, 2013).

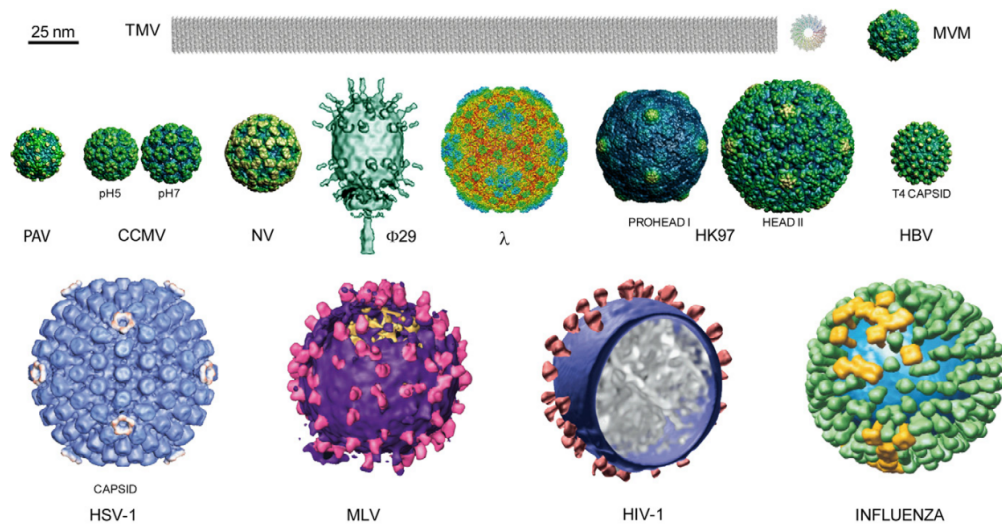


Figure 1.1 Some representative viruses shown in the same scale. From left to right and top to bottom: Helical virus: TMV (side and top view). Noneveloped viruses: MVM (T=1); PaV (T=3); CCMV (T=3, native and swollen forms); NV (T = 3); Phi29 (T = 7, prolate); λ (T = 7, tail not shown); HK97 (T = 7, immature prohead-I and mature head-II). Icosahedral capsids of enveloped viruses: HBV capsid (T = 4); HSV-1 capsid (T = 16). Enveloped viruses: MLV; HIV-1 (in which the virion and internal cone-shaped capsid have been sectioned to show the distribution of other internal macromolecules); influenza virus (Mateu, 2013a).

In 1935, Wendell M. Stanley for the first time isolated tobacco mosaic virus (TMV) from tobacco leaves from when information on size, shape, and properties started to be known (Kay, 1986). Virus particles vary remarkably in size, shape, molecular composition and structural organization (Figure 1.1). Nonenveloped viruses have the simplest composition of a protein shell or capsid which contains a genome of DNA or RNA while in some more complex virus particles, the composition inside the capsid may also contain proteins and/or other macromolecules. External protein structures like portals, tails, fibers, etc. may also be attached to the virions for specific functions (Rossmann, Rao, & Venigalla, 2012). Enveloped viruses are surrounded by a lipid bilayer embedded with proteins, the envelopes are typically derived from portions of the host cell membranes (phospholipids and proteins), but include some viral glycoproteins. They may help viruses avoid the host immune system. Glycoproteins on the surface of the envelope serve to identify and bind to receptor sites on the host's membrane. The viral envelope then fuses with the host's membrane, allowing the capsid and viral genome to enter and infect the host (Rossmann et al., 2012).

Although virus structures are hugely diverse, their capsids can be generally described as hollow symmetric oligomers or multimers made of many copies of one or a few different types of capsid protein (CP) subunits (Mateu, 2013b). Among all the shapes of virus capsids, around half of the virus families have an icosahedral arrangement. The simplest icosahedral virus capsids are made of only 60 identical CP subunits. Satellite tobacco necrosis virus (STNV), poliovirus, bacteriophage PhiX174 and parvoviruses such as the minute virus of mice (MVM) are all examples of icosahedral viruses. Around 10% of viruses adopt a helical shape as seen for tobacco mosaic virus, leaving the rest to be composed of less simple arrangements of CP subunits (Mateu, 2013b).

1.1.2. Structural study of viruses using X-ray crystallography

In a high concentration of virus solution, a high degree of order is suggested by birefringence in a polarizing light microscope. This method was used by Bernal and Fankuchen on TMV suspensions to generate an X-ray diffraction pattern from which they established the size and shape of the first detailed description of any virus structure (Bernal & Fankuchen, 1941). Taking advantage of Svedberg's centrifugation techniques which made the purification procedures simple, and the possibility of extracting milligrams of virus from a kilogram of leaf material, the crystalline viruses studied with X-ray diffraction in the early days were mainly RNA plant viruses (Crowfoot, 1945; Huxley & Zubay, 1960; Karlson & Luscher, 1959; Purcell, 1956). These discoveries were even earlier than the structures of 17 kDa myoglobin and 64 kDa haemoglobin structures determined in 1957, as the first protein structures (Kendrew et al., 1958; Perutz, Rossmann, Cullis, Muirhead, & North, 1960), and the molecular weight of the particles were at least 100 times larger than haemoglobin. These crystalline virus structures determined in the early times were mostly spherical in shape with diameters larger than 200 Å with a protein capsid enclosing nucleic acid and the structures were determined at much lower resolution than those of myoglobin and haemoglobin. With the discovery of the relationship between DNA sequence and the protein product it encodes (Watson & Crick, 1953), Watson and Crick then made a prediction that due to the impossibility of the genome to have enough capacity to encode sufficient protein to form the outer capsid of a virus, the capsid of viruses must be formed by a symmetrical arrangement of repeated identical proteins (Crick & Watson, 1956).

In 1978, the first atomic viral structure was published for tomato bushy stunt virus at 2.9 Å resolution (S. C. Harrison, Olson, Schutt, Winkler, & Bricogne, 1978) To now,

there are nearly 8000 viral related protein structures from around 500 different viruses in the PDB among which more than 80% of these structures are only viruses, and a large majority (89%) of these structures were determined with X-ray crystallography. The quantity structures deposited into the archive has been growing through the years and the resolution of structures solved with X-ray crystallography has remained at around 2.5 Å (Bassetto, Massarotti, Coluccia, & Brancale, 2016). An X-ray crystal structure of the VLP formed by the coat protein (CP) of the Satellite Tobacco Necrosis Virus (STNV) has been solved and refined at 1.4 Å resolution, representing a virus structure determined at a super high resolution (Lane et al., 2011).

1.1.3. Application of virus structures

Structural biology on viruses can be applicable, especially in the antiviral drug discovery process. One of the most striking examples is the HIV-1 protease and reverse-transcriptase inhibitor discovery, which led to the first selective HIV-1 protease inhibitor developed with the use of X-ray crystallography (Szelke et al., 1982). The new approach of structure-based antigen design was predicted to become an important path for vaccine innovation in the future, with several examples of new vaccine antigens delivered in the past few years based on structure-based design, or structural vaccinology (Dormitzer, Grandi, & Rappuoli, 2012).

1.2. Structural Studies of Bluetongue Virus VP5

1.2.1. Brief Introduction of Bluetongue Disease

Bluetongue (BT) is an arthropod-transmitted disease of ruminates, especially sheep and cattle (Saegerman, Berkvens, & Mellor, 2008; A. J. Wilson & Mellor, 2009). The majority of infected animals develop no detectable clinical signs but in some cases

symptoms may include lameness, depression, fever, oedema of the lips, tongue and head, coronitis, conjunctivitis, excessive salivation, nasal discharge, hyperaemia and death (A. J. Wilson & Mellor, 2009) (Figure 1.2). The blue tongue after which the disease was named is only seen in some rare and more serious clinical cases. After recovery, animals may also carry on with some long-lasting secondary effects like reductions in milk yield and weight gain, severe wool break and temporary infertility. Normally, severe cases can only be seen in certain breeds of sheep—European fine wool and mutton breeds, and some wild ruminants like North American white tailed deer (Deer & Vosdingh, n.d.; N J Maclachlan, Drew, Darpel, & Worwa, 2009). Bluetongue virus (BTV) is the viral pathogen of bluetongue disease, which is the type species of *Orbivirus* genus in the family *Reoviridae*. Ten segments of double-stranded RNA are contained in a shell of three concentric layers of subcore, core and capsid. The individual proteins forming the shell will be further described in the later contents.

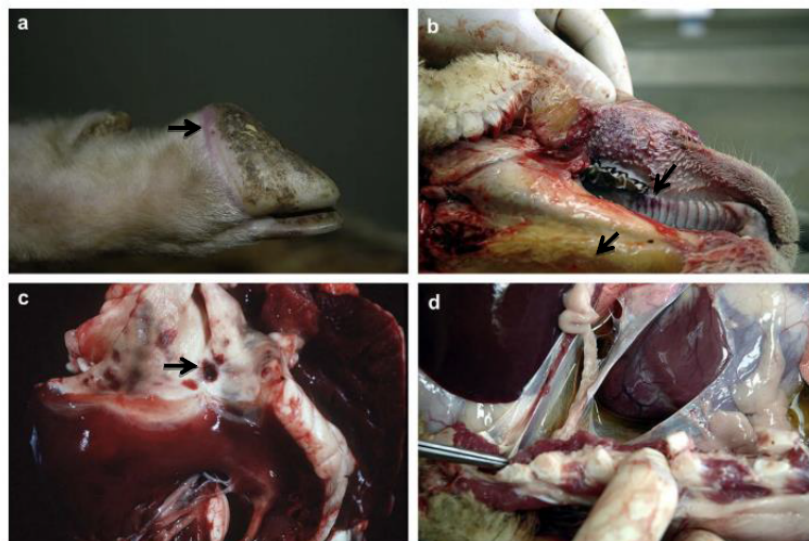


Figure 1.2 Symptoms of sheep with bluetongue. (A) Distal limb: coronitis (arrowhead). (B) Head: submandibular oedema (arrowhead) and hemorrhage and ulceration (arrow) of the hard palate and dental pad. (C) Heart: subintimal hemorrhages (arrowhead) in the pulmonary artery. (D) Pericardial space: pericardial effusion (Maclachlan N.J. *et al* 2009).

Due to its serious impact on animal health, BT is listed as a “notifiable disease” by the World Organization for Animal Health. In 1905, Spreull published the first comprehensive clinical profile of bluetongue disease. During the century after this, the world distribution of this disease is a broad band between within the 40°N and 35°S while a dramatic northwards extension of the occurrences happened more recently (A. Wilson & Mellor, 2009). The reason for the extension may involve several different mechanisms including the movement of infected livestock or the migration of the *Culicoides* virus carriers. The distribution of BTV serotypes is not uniform globally and undergoes drastic changes possibly due to the change of global climate (N James Maclachlan, 2010; Saegerman et al., 2008).

1.2.2. Serotypes of BTV

To date, there are 27 serotypes of the virus identified on all continents except for Antarctica (Jenckel et al., 2015). In the United States, BTV was initially isolated from sheep during a 1952 outbreak (Hardy & Price, 1952). Over the next 30 years, BTV-2, -10, -11, -13, and -17 were gradually isolated from livestock and wildlife in the US, characterizing the endemic BTV serotypes (Gibbs E, Greiner E, Taylor W, Barber T, 1983). From 1999, the number of BTV serotypes noted in the US has been enlarged after the detection of 10 historically non-endemic serotypes (1, 3, 5, 6, 9, 12, 14, 19, 22, and 24) in the south-eastern United States (Gibbs EPJ, Tabachnick WJ, Holt TJ, 2008).

In August 2006, an unprecedented outbreak caused by a BTV serotype 8 strain (BTV-8_{net2006}) occurred in north-western Europe (Sperlova & Zendulkova, 2011). This single BTV outbreak was the most economically damaging ever. This particular BTV-8 strain is also epizootically and pathologically unique. Following the relatively

limited initial outbreak in 2006 with which only 2000 holdings in the Netherlands, Belgium, Luxemburg, Germany and France were infected, the virus re-emerged after the winter period due to vector survival, which is also called overwintering, which was previously considered unlikely (N. James Maclachlan & Mayo, 2013a). Trans-placental transmission of virus may also have contributed to the overwintering. By the end of 2007 nearly 60 000 holdings had been infected. The infection areas ranged from Sweden to Italy and from the UK to the Czech Republic and Hungary (Saegerman et al., 2008). The virus was spread by *Culicoides obsoletus* and *C. pulicaris*, which were previously not known to be very effective BTV vectors in the field (A. Wilson & Mellor, 2009). In addition to BTV-8, BTV-6 was detected in Netherlands and Germany while BTV-11 was found in Belgium (De Clercq et al., 2009).

Among the 27 serotypes of the virus identified (Jenckel et al., 2015), lifelong protection can be generated to a serotype due to infection, although there is no apparent cross-serotype protection. Transmission is predominantly through a *Culicoides* vector, although other vectors and routes have also been reported which are considered of relatively minor importance.

Bluetongue is considered to be a disease of severe economic impact with a global estimation of US\$ 3 billion. When it comes to the BTV-8 outbreaks, Wilson and Mellor stated that the “BTV-8 epidemic in Northern Europe has probably caused greater economic damage than any previous single-serotype Bluetongue outbreak” (A. J. Wilson & Mellor, 2009). Losses can be classified as losses in direct production such as reduced milk yield or increased mortality, weight loss, reduced fertility rate, abortion, reduced meat production efficiency and death (Sperlova & Zendulkova, 2011). Indirect losses of expenditure and lost revenue include costs of vaccines or lost revenue, such as through trade restrictions limiting access to higher value markets.

1.2.3. Transmission and Controlling of BTV

BTV is transmitted among the ruminant hosts by certain species of biting midges of the genus *Culicoides* which are small (< 3mm) haematophagous insects that exist in all inhabited continent in the world. These insects breed in a wide variety of semi-aquatic sites including tree holes, rotting vegetation, pond margins, damp soils and certain sorts of herbivore dung. Out of more than 1000 species of *Culicoides* known worldwide, only a few of them have been incriminated as vectors of BTV. Thirty species of *Culicoides* biting midges that play a greater or lesser role in the transmission of bluetongue disease in the pantropical regions of the world were listed in the research of R. Meiswinkel (Meiswinkel, Gomulski, Delécolle, Goffredo, & Gasperi, 2004), including *C. imicola*, the principal vector in Africa, the Middle East, much of southeast Asia and parts of Southern Europe, *C. sonorensis*, the principal vector in North America and *C. brevitarsis*, the principal vector in Australia.

Current and potential strategies for controlling BTV infection of wild and domestic ruminants are either through protecting the animal hosts from infection or by elimination of the midge vector. The former animal-based control strategy mainly includes vaccination and livestock housing, movement restrictions and culling (N. James Maclachlan & Mayo, 2013a).

Vaccination is currently central to the response of most at-risk countries to any BT outbreak. National policies of vaccinating livestock based on the geographic distribution of BTV serotypes have been implemented by authorities in affected European countries to limit the losses and prevent spread of BTV (Noad & Roy, 2009). Vaccines of BTV have been developed which include whole killed viruses, virus like particles (VLPs), modified live virus (MLV), and vaccines developed from recombinant vaccinia, canarypox or capripox virus vectors (Noad & Roy,

2009)(Savini, MacLachlan, Sanchez-Vizcaino, & Zientara, 2008) (Di Emidio et al., 2004). With benefits and disadvantages of each type of vaccine, only some inactivated virus vaccines and some MLPs can be used in the European Union approved national disease control programs (Savini et al., 2008) (Zientara, MacLachlan, Calistri, Sanchez-Vizcaino, & Savini, 2010).

Except for controlling infection from the ruminant hosts side, there are also vector-based strategies to reduce or eradicate the population of mature or pre-mature midge vectors. Controlling the immature stage of *Culicoides* midge vectors can be achieved by either reducing larval habitat and breeding sites such as the drainage of marshlands, waste-water lagoons, or standing pools of water, or using organochlorine, organophosphate (OP), or pyrethroid insecticides (Carpenter, Mellor, & Torr, 2008) (N. James MacLachlan & Mayo, 2013b). To target mature stage of *Culicoides* midges, insecticides can be used to reduce populations of mature vectors and thereby reduce the risk of BTV transmission. Repellents and attractants have also been evaluated as “decoys” to lure *Culicoides* midges to reduce the biting rate among ruminant hosts.

1.3. Overview of Structural Studies on Bluetongue Virus and Protein Functions in the Virus Life Cycle

1.3.1. Proteins of BTV

The 10 linear dsRNA pieces (Seg-1—Seg-10) that the bluetongue virus contains encode 7 structural proteins (VP1-VP7) that form a complex structure of a core (VP3+VP7) (Grimes et al., 1998), subcore (VP7) and capsid (VP2 and VP5) layer. These layers encapsidate the genome, as well as the viral polymerase (VP1) (Matsuo & Roy, 2011), capping enzyme (VP4) (Sutton, Grimes, Stuart, & Roy, 2007) and the

helicase (VP6) (Kar & Roy, 2003). In addition to these seven structural proteins, BTV also synthesizes 5 non-structural proteins (NS1, NS2, NS3, NS3A, and NS4) in infected cells, each of which is involved in various stages of virus replication and assembly (Beaton, Rodriguez, Reddy, & Roy, 2002; Boyce, Celma, & Roy, 2012; Feenstra, van Gennip, Maris-Veldhuis, Verheij, & van Rijn, 2014; Horscroft & Roy, 2000; Mumtsidu, Makhov, Roessle, Bathke, & Tucker, 2007; Owens, Limn, & Roy, 2004; Ratnien et al., 2016).

1.3.2. The structural proteins of BTV and their functions

Cell entry utilizes clathrin-mediated endocytosis and pH-dependent penetration which involves the dissociation of the outer-capsid, composed of VP2 and VP5, and then leads to the release of the core in the cytosol (Forzan, Marsh, & Roy, 2007). The BTV virion contains an outer layer of 60 VP2 trimers and 120 VP5 trimers. The VP2 trimers form 'triskelion' motifs with three interlocked spirals on the outer layer which is responsible for receptor binding, hemagglutination and eliciting serotype-specific neutralizing antibodies (S. S. Hassan & Roy, 1999). A recent cryo-EM structural study on the intact BTV virions with a direct electron detector operated at super-resolution and reconstructed by single-particle analysis gave a structure of 3.5-Å-resolution at high pH (Figure 1.3) and a lower resolution structure filtered to 15 Å at acidic pH (Figure 1.4). The low pH structure was with detached VP2 and with VP5 undergoing large-scale conformation changes leading to outward protrusion (X. Zhang et al., 2016). This study also proposed a model for cell entry which involves the detachment of VP2 when BTV is exposed to the low pH and a dramatic conformation change in VP5.

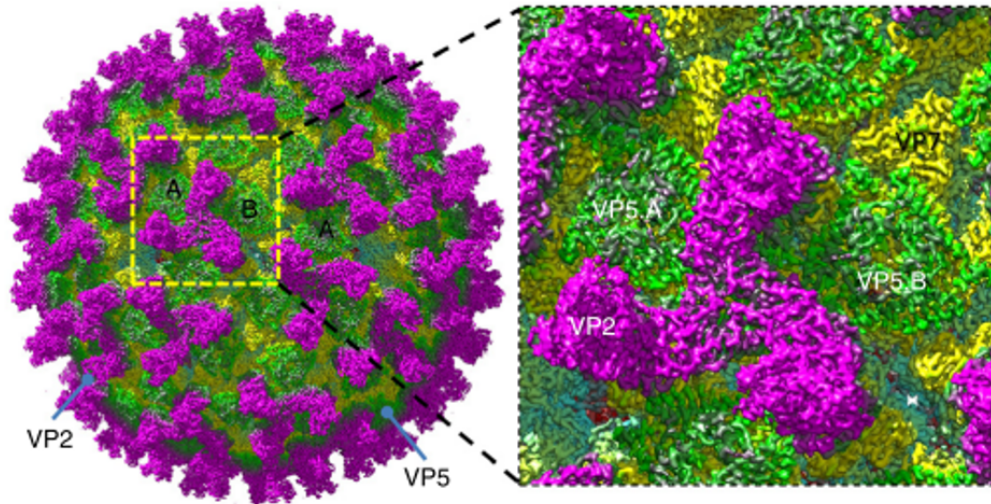


Figure 1.3 Cryo-EM density map of the BTV virion. The electron density map is shown as radially colored surface representation, and a close-up view of the boxed area containing an asymmetric unit. (X. Zhang et al., 2016).

Cell entry of BTV involves removal of the outer capsid, leading to release of the core into the cytoplasm of the target cell. The bluetongue virus core, the transcriptionally active compartment of the virus, is 700 Å in diameter and is composed of two principal structural proteins. 780 copies of the VP7(T=13) protein (Mr=38kDa) and 120 copies of the VP3(T=2) protein (Mr=100kDa) form the subcore which is surrounded by 260 copies of VP7 trimers. The structure of the core has been determined by X-ray crystallography at a resolution approaching 3.5 Å which was the largest molecular structure determined in such detail at the time (Grimes et al., 1998) (Figure 1.5). The subcore of VP3 has 60 subunits arranged with icosahedral symmetry. Each subunit consists of 2 VP3s which are in different conformation. In the arrangement of VP7 on the core particles the trimers can occupy five different positions represented by P, Q, R, S, and T. The VP7 forms the outer layer of the core outside of the VP3 subcore. The latter can retain its particle structure if the outer VP7 core layer is missing (for

example as expressed by recombinant baculovirus in insect cells) (Huismans, van Dijk, & Els, 1987). This means that the information delivered by the amino acid sequence of VP3 must determine not only the fold of the protein itself, but also the self-assembly of the subcore which can further be used as the scaffold for VP7 to attach and subsequently the outer capsid proteins VP2 and VP5 to assemble onto. VP3 not only

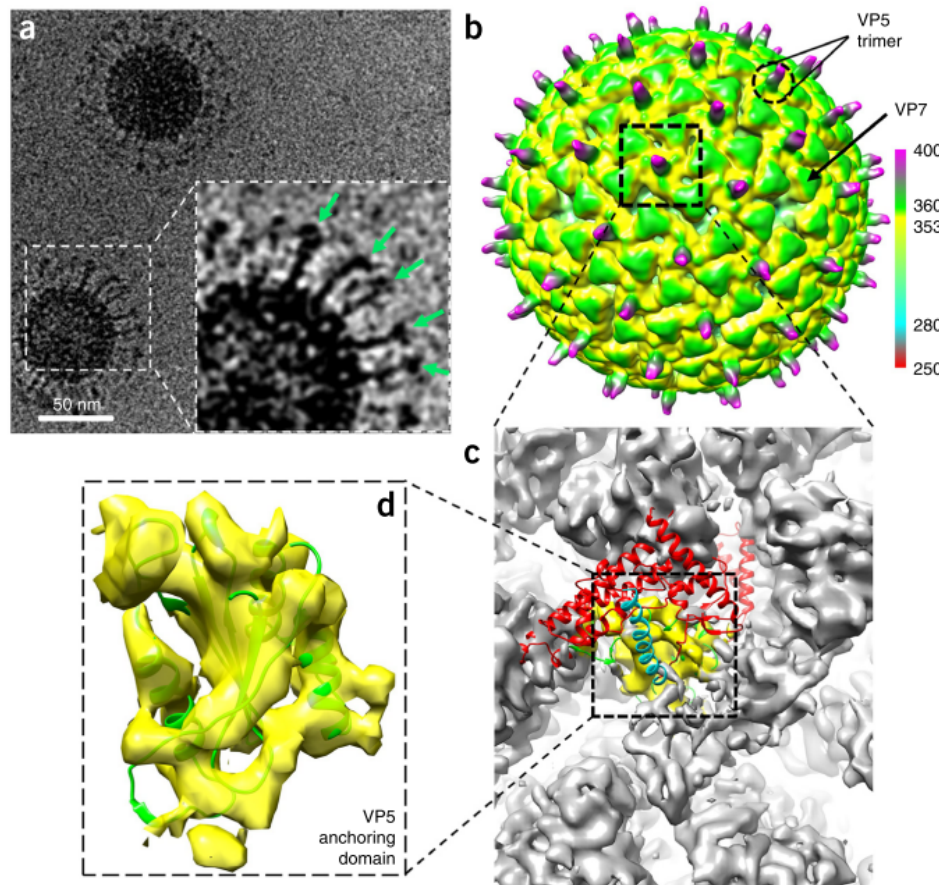


Figure 1.4 Structure of VP5 at low pH.(A) Cryo-EM image of the BTV virion at pH 3.4, showing a dramatic conformational change of VP5. (B) Cryo-EM construction of the BTV virion at acidic pH, filtered to 15-Å resolution and displayed at threshold of 1.0σ, showing that VP2 is detached, and VP5 has undergone large-scale conformation changes. (C) Close-up view of the region within the dashed box in b. The ribbon model of the VP5 monomer at high pH is fitted on the VP5 density (gray and yellow) in the reconstruction of the virion at low pH. (D) Close-up view of the yellow region in C. (X. Zhang et al., 2016).

determines the interaction with the external proteins, it also interacts with the inner components of the particles, including the transcriptase complexes composed of VP1, VP4 and VP6 and the genome (Gouet et al., 1999). How the core is assembled remains a question in virus-infected cells. What has been studied is that viral inclusion bodies (VIBs), which are predominantly driven by viral-encoded NS2, act as the assembly sites for core components and the intermediate core could also successfully recruit VP1 and VP4 but not VP6 (Kar, Bhattacharya, & Roy, 2007). The recruitment of ssRNAs is essential to drive the functional core assembly (Lourenco & Roy, 2011).

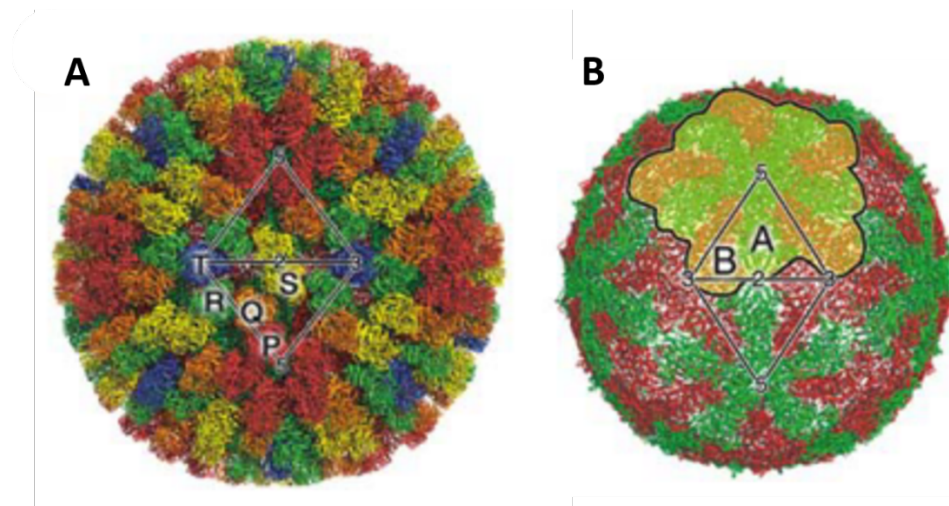


Figure 1.5 The core particle of BTV-1.(A) The icosahedral asymmetric unit (which is a triangular area defined as marked by the symmetry axes of the icosahedron) contains 13 copies of VP7(T13) arranged as 5 trimers, P, Q, R, S and T, coloured red, orange, green, yellow and blue, respectively. (B) The VP3(T2) scaffold. The icosahedrally unique molecules A and B are coloured in green and red, respectively (Grimes et al., 1998).

BTV and other members of the *Reoviridae* differ from other RNA viruses during virus disassembly in that their transcription complexes remain within the core, instead of being released from the particles (Eberhard-karls, Mainou, & Dermody, 2016). Only the single-stranded RNA (ssRNA) transcripts (mRNAs) are released from the core.

The 5-fold pores expanded when the core crystals were soaked in transcription buffer, giving an indication that during transcription the mRNAs exit from these pores, while low molecular weight metabolites such as nucleotides use different channels for entry (Diprose et al., 2002). The BTV core has 3 minor proteins, VP1, VP4, and VP6, that are responsible for synthesising the ‘capped’ and methylated transcripts of each dsRNA segment that are released from the transcribing core into the cytoplasm. Among these three minor proteins, the largest one, VP1 (149.5 kDa), is the RNA-dependent RNA polymerase (RdRp). Although there is no structural information available for VP1 currently, manual alignment of the BTV protein against Reovirus and other viral RdRps generated a model with three domains similar to other RdRps: An N-terminal domain, a C-terminal domain, and a central polymerase domain (Wehrfritz, Boyce, Mirza, & Polly, 2007). In these RdRps, the active site is at the centre of the palm domain. Experiments showed that in a single *in vitro* reaction, VP1 could synthesize the 10 dsRNAs simultaneously from BTV core-derived ssRNA templates and synthesis of dsRNAs from capped ssRNA templates was significantly higher than that from uncapped ssRNA templates (Matsuo & Roy, 2011).

The cap (methylguanosine connected to the first nucleoside) stabilizes the mRNA and allows efficient translation. While many viruses use the cellular capping machinery to modify their mRNAs, BTV and some other double-stranded RNA viruses encode their own enzyme. In BTV this is the second largest minor protein VP4 (76.4 kDa) which caps viral mRNA molecules before they are released from the transcribing viral core (Mertens, 2004). This capping process requires four distinct enzymes which are: an RNA triphosphatase (RTase) that hydrolyses the 5'-triphosphate terminus of the mRNA to a diphosphate; a guanylyltransferase (GTase) that caps the diphosphate terminus with GMP via a 5'-5' triphosphate linkage; a guanine-N7-Methyltransferase

(N7MTase) that adds a methyl group to the N7 position of the blocking guanosine; and a nucleoside-2'-O-methyltransferase (2'OMTase). In BTV, these four enzymatic reactions are all performed by VP4. The crystal structure of VP4 published by Sutton (Sutton et al., 2007) revealed an elongated modular architecture that provides a scaffold for an assemblage of active sites with an overall linear layout (Figure 1.6).

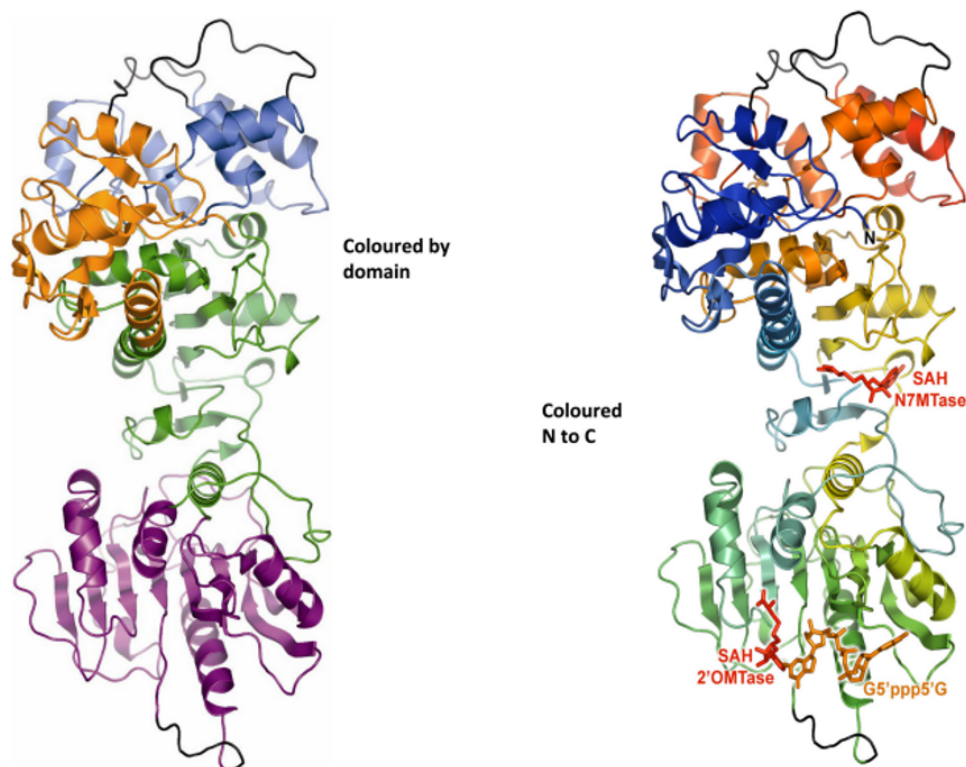


Figure 1.6 Cartoon representation of the X-ray structure of the VP4 molecule. Ribbon diagram of the VP4 structure showing four discrete domains that are arranged in a linear fashion (left). On the right is the ribbon diagram of the VP4 structure coloured from blue at the N terminus to red at the C terminus. The cartoon on the right also shows the ligand binding sites (Sutton et al., 2007).

The smallest BTV core protein, VP6, is an RNA-dependent ATPase with helicase activities which possesses nucleoside triphosphatase, RNA binding and helicase activities that has been confirmed by *in vitro* assay (Martinez-costas, Sutton, Monastyrskaya, & Roy, 1997). It binds both single-stranded and double-stranded RNA since it is rich in lysine and arginine residues (P Roy, Adachi, Urakawa, Booth, & Thomas, 1990). This gives evidence for VP6 being a nucleic acid-unwinding protein, which binds a single-stranded region for initial attachment and could also bind to and initiate helicase activity on a duplex RNA. A number of functional motifs can be recognized in the BTV VP6 protein when it is compared with other helicases such as putative sequences for RNA binding, ATP binding, and helicase function as well as a Walker A motif, and the protein is in hexameric state and forms ring-like structures when complexed with ssRNA or dsRNA (Kar & Roy, 2003). When site-specific mutagenesis was introduced into these putative functional sites, the functional activity of each VP6 mutant derivative was reduced or lost in comparison to the wild-type VP6. (Kar & Roy, 2003).

1.3.3. The non-structural proteins of BTV and their functions

NS1 is the 64 kDa non-structural protein that highly accumulates in the cytoplasm and forms into tubular structures after orbivirus infection which no homologue has been found in other genera of the *Reoviridae* (Hewat, Booth, Wade, & Roy, 1992). Being the most highly expressed BTV protein before completion of the replication cycle, knowledge of the functions of NS1 remains limited to a certain role in regulating the mechanism of virus release from cells (Owens et al., 2004). Using an mRNA reporter system, work done in 2012 identified NS1 as a positive regulator of viral protein synthesis which upregulates the synthesis of viral proteins and creates a positive

feedback loop of NS1 expression followed by an increase in the expression of all the viral proteins (Boyce et al., 2012). The replacement of cellular protein synthesis for viral protein synthesis during infection can be attributed to the efficient translation of viral reporter mRNAs among cellular mRNAs (Boyce et al., 2012).

NS2 might be involved in some way in selection and packaging of the viral RNAs. Evidence includes the observation that although there are some weak preference for which genomic segments the NS2 protein binds to, no specific RNA binding property of NS2 was identified (Zhao, Thomas, Bremer, & Roy, 1994). The sequences of regions 2–11, 153–166 and 274–286 are reported to be important for RNA binding in a more recent study (Fillmore, Lin, & Li, 2002). Other supporting evidence is that the NS2 has been shown to be able to hydrolyse nucleotide triphosphates to nucleotide monophosphates (Horscroft & Roy, 2000). NS2 plays a key role in the formation of viral inclusion bodies (VIBs) where the assembly of new virions takes place (Brookes, Hyatt, & Eaton, 1993). The x-ray structure of the N-terminal domain of BTV NS2 was determined to 2.4 Å resolution (Figure 1.7) and shows a polymeric structure mediated by two monomer–monomer interfaces (Mumtsidu et al., 2007).

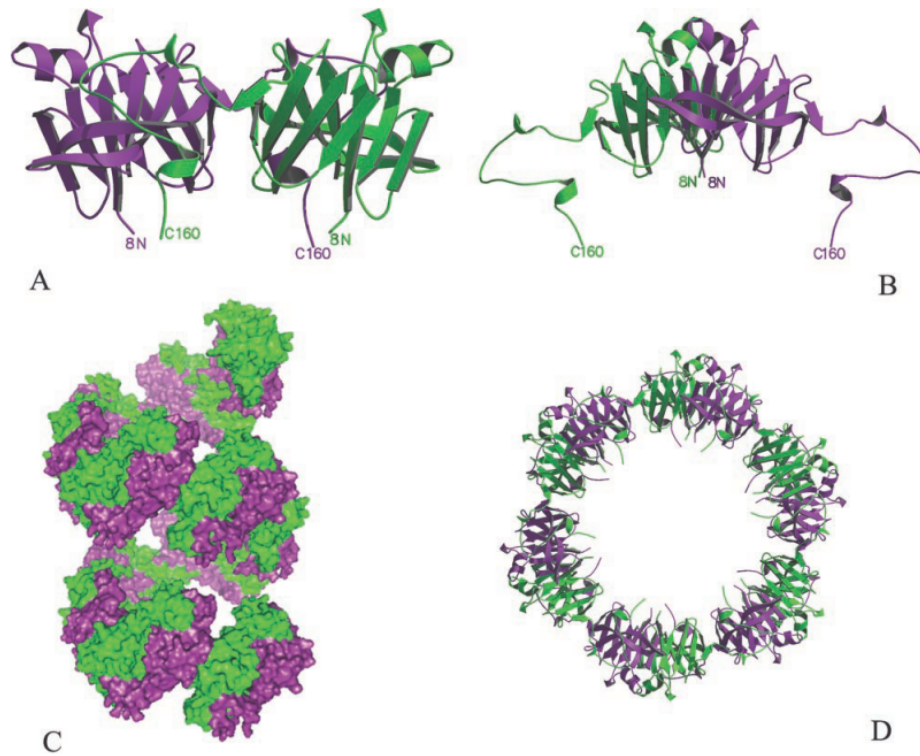


Figure 1.7 Ribbon representation of the NS2N homodimeric configurations and overview of the crystal packing. (A) the largest interface between monomers. (B) the extensive sheet interface (C) homodimeric associations give rise to an infinite spiral in the crystal with the pitch equal with the crystallographic c axis. (D) crystal packing viewed along the 6₅ crystallographic axis, which is perpendicular to the plane of the figure (Polly Roy, 2009).

NS3 proteins are encoded by BTV segment 10 (Seg-10). There are two isoforms, NS3 and NS3a, the latter is translated from a second in-frame initiation codon and lacks the first N-terminal 13 amino acid residues of NS3 (Bansal, Stokes, Bansal, Bishop, & Roy, 1998). They are viroporin-like membrane proteins as well as glycoproteins that localize in the Golgi apparatus and at the plasma membrane of infected cells where they mediate BTV egress through budding and impede the *in vitro* synthesis of type I interferon in cells. NS3/ NS3a is composed of a long N-terminal and short C-terminal cytoplasmic tails linked by two transmembrane domains and a short extracellular

domain with a single asparagine glycosylation site (Bansal et al., 1998). A very interesting feature of BTV infection is that in mammalian cells, although most of the mature virions are released by cell lysis after infection, some virus particles are released from infected cells by local extrusion and budding from the plasma membrane while in insect cells, release of these viruses is almost entirely non-lytic which is a process proved to be correlated with high level expression of NS3 in invertebrate cells infected with BTV (Guirakhoo, Catalan, & Monath, 1995). BTV NS3 is suggested to interact with a cellular protein component, the tumour susceptibility gene 101 protein (Tsg101), an ESCRT-I complex which is normally involved in the last stages of membrane fission (Wirblich, Bhattacharya, & Roy, 2006). The C-terminal domain of NS3 is also shown to interact with VP2, the outermost protein of the fully assembled virus particle (Beaton et al., 2002). These data suggest that NS3 may form a bridge between the virus particle and the host cell budding machinery. Another recent study shows that without expression of NS3/NS3a proteins in both mammalian and insect cells could result in a strongly reduced induction of cytopathogenic effect (CPE) and a reduced virus release which suggests that a NS3/NS3a knockout virus could be safe *in vivo* thus may be a potential BTV vaccine candidate. Tests performed on sheep confirmed the NS3/NS3a knockout bluetongue vaccine candidates are avirulent and protect against virulent BTV challenge (Feenstra et al., 2014).

NS4 is a small protein (77aa) with nuclear localization function encoded by an ORF in genome segment 9 overlapping a larger ORF encoding VP6. A recent study showed that BTV NS4 is an interferon (IFN) antagonist and a key determinant of viral virulence (Ratinier et al., 2016). It is shown that NS4 not only confers a replication advantage on the virus in IFN-competent primary sheep endothelial cells and immortalized cell lines, it also downregulates the activities of a variety of promoters

including the cytomegalovirus immediate-early promoter, the IFN- β promoter, and a promoter containing interferon-stimulated response elements (ISRE). An increased synthesis of type I IFN is detected in an NS4 deletion mutant compared to cells infected with wild-type BTV (Ratinier et al., 2016).

1.3.4. Non-enveloped viruses (especially BTV) cell entry

Viruses need to overcome several barriers to deliver their nucleic acid into the cytosol or get to the nucleus to perform their own replication by hijacking machinery from their host cells and cause infection. The cellular membrane forms a very important barrier for extracellular viruses to get through. Viruses have developed different strategies to breach this barrier. Enveloped viruses achieve cell entry through a well characterized and relatively straightforward way, via direct fusion of the outer viral membrane with a cellular membrane, with the help of the encoded hydrophobic peptides to promote fusion resulting in a transfer of the nucleocapsid to the cytoplasm (Poranen, Daugelavičius, & Bamford, 2002). To take the influenza virus as an example, the viruses are first taken up into endosomes where the low-pH environment triggers a conformational change which exposes a hydrophobic viral peptide and initiates the fusion reaction between the viral and endosomal membranes (Jahn & Sudhof, 1999).

However, in the case of non-enveloped viruses which lack a lipid bilayer, the membrane penetration mechanism is fundamentally different from that of enveloped viruses. During the beginning of the infection, the non-enveloped viruses are first taken in by receptor-mediated endocytosis, often facilitated by a clathrin- or caveolae-dependent process, the two major endocytosis pathways (Marsh & Helenius, 2006). Examples have been reported such as for canine parvovirus which enters host cells via clathrin-dependent endocytosis by binding to the transferrin receptor and are

transported to recycling endosomes (Parker, Murphy, Wang, O'Brien, & Parrish, 2001). Also adenovirus reaches the endosomes by binding to the integrin $\alpha\beta 5$ receptor and internalizes (Pickles et al., 1998). Papillomavirus also traffics to endosomes via a clathrin-dependent pathway while the specific host cell receptor remains uncertain (Day, Lowy, & Schiller, 2003). BTV also exploits clathrin-mediated endocytosis for productive infection, suggested by thin-section EM studies previously showing the presence of BTV particles in coated pits (Eaton, Hyatt, & Brookes, 1990). Also adding in Chlorpromazine (CPZ) which inhibits the formation of clathrin vesicles resulted in the inhibition of BTV replication (Forzan et al., 2007). Recent studies on the mechanisms of virus entry have shown that the different endocytic processes do not act in total isolation, rather there are a number of interconnections among endocytic components and routes, resulting in a complex network of internalisation pathways (Kalia & Jameel, 2011). After internalization, the viruses are then trafficked to the membrane penetration site.

After being trafficked to the penetration site, non-enveloped viruses may engage various cellular factors such as receptors, low pH, proteases, or chaperones that trigger truncation of viral proteins or conformational changes to viral proteins. This leads to the formation of activated viral intermediates that initiate the membrane transport process by forming or exposing a hydrophobic moiety or membrane lytic factor. There are three major classes of structural features that disrupts the membrane: amphipathic α -helices, myristoyl groups, and lipid-remodeling enzymatic domains while mechanistically three categories of forms of membrane disruption were observed as transient modification of the cellular membrane, pore formation, and total disruption of the limiting membrane (Johnson, 2010).

To take poliovirus as an example, the interaction between the cellular receptor PVR and poliovirus is shown to initiate the penetration process by causing an irreversible conformational change which exposes VP4 and the N terminus of VP1 which then binds and inserts into a membrane (Fricks & Hogle, 1990). In another case, during the early stage of infection by reovirus (Figure 1.8), the virus is exposed to the acid-dependent proteases within the endolysosomal system. The outer protective protein $\sigma 3$ is cleaved by the proteases cathepsins B and L (Ebert, Deussing, Peters, & Dermody, 2002). This cleavage converts the native viral particle to the infectious subviral particle (ISVP) and then being converted to a metastable particle which initiates membrane penetration by interacting with a yet unidentified cellular factor. In the form of ISVP*, the N-terminal myristoyl moiety ($\mu 1N$) is exposed which enables an interaction with the lipid bilayer (Chandran, Farsetta, & Nibert, 2002). A similar method is adopted for papillomavirus entry. The N terminus of the L2 protein of papillomavirus is cleaved by a protease, furin, in the trans-Golgi causing a conformational change of the C terminus and thereby priming the virus for membrane penetration since the C terminus of L2 has membrane disrupting activity (Kämper et al., 2006). Another relatively well studied virus is rotavirus, in which the proteolytic activation, trypsin-mediated cleavage in this case, is also essential in the penetration process. The outer capsid protein VP4 of rotavirus is cleaved to generate VP8* and VP5*, with VP5* being shown to have the membrane-disrupting function (Denisova et al., 1999).

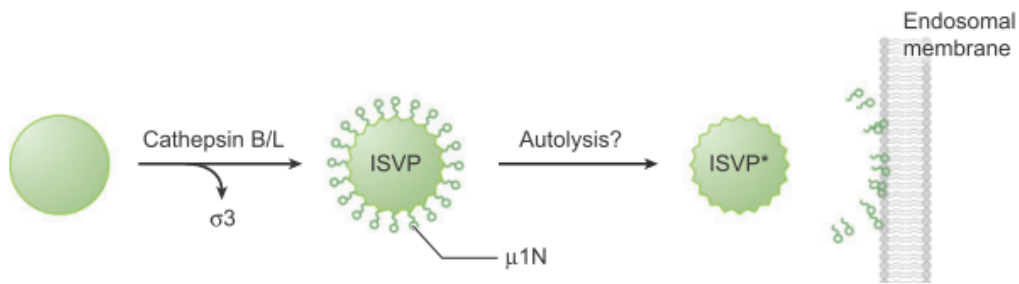


Figure 1.8 Mechanisms that initiate the membrane penetration process. (Tsai, 2007).

Other than the proteases triggered viral conformational changes described above, low pH represents another cellular condition that can trigger the penetration of non-enveloped viruses, similar to the enveloped viruses in which the fusion between the viral and cellular membranes is induced by low pH. Through evolution, viruses developed mechanisms to sense and detect pH variation in the acidifying vesicles to trigger conformational changes to proteins or regulate the endosomal enzyme reactions which promote cytosolic release. Some particular amino acids in the virus proteins can be used to sense pH. The ionisable side chain of histidine (His) has a pKa of 6.4 which is about the pH level inside early endosomes. As in the case of adenovirus, the release of the internal protein VI with the ability to disrupt the endosomal membrane is facilitated by the destabilization of the adenovirus capsid by endosomal low-pH environment (Wiethoff, Wodrich, Gerace, & Nemerow, 2005). The protein VI is expressed as a 250 amino acid long pre-protein and protease cleavage generates the 206 amino acid mature protein by removing both N- and C-termini. The N-terminal domain of protein VI, which forms an amphipathic α -helix, is required for the lysis of the endosome and this domain is highly conserved among adenoviral species. Both the precursor and the mature form of protein VI have the membrane disruption ability in a pH-independent way (Wiethoff et al., 2005). Mutations in the α -helix domain are

shown to lower its membrane disruption activity (Moyer, Wiethoff, Maier, Smith, & Nemerow, 2011). Another example is parvoviruses, the VP1 N-termini of the virus has a secretory PLA₂ homology domain which is a lipolytic enzyme that destroys the membrane to allow the viruses to escape from the lysosome. The PLA₂ homology domain is initially located within the native capsid but exposed after low pH treatment (Ros, Gerber, & Kempf, 2006).

The feature of an α -helix domain exists in more than adenovirus protein VI described above, it can also be found in the 21 residues from N-terminal of the γ peptide of Flock house virus (FHV) which is capable of mediating liposome lysis. The helical conformational change which results in membrane lysis was discovered by mixing with model membranes (Janshoff, Bong, Steinem, Johnson, & Ghadiri, 1999). The N-terminal region of Poliovirus (PV) capsid protein VP1 also contains an amphipathic α -helix domain which is first buried but gets externalized by a receptor-binding induced conversion (Moyer & Nemerow, 2011; Ren et al., 2015).

1.3.5. Structure features of BTV related to cell entry

The BTV VP5 protein also has a pH-dependent fusogenic activity when expressed on the cell surface (Forzan, Wirblich, & Roy, 2004). The relevance between the acidic pH and BTV entry was initially suggested by showing that the presence of ammonium chloride inhibited BTV replication (Hyatt, Eaton, & Brookes, 1989). Additional evidence from an experiment showing that reduction of virus infection is strictly correlated with bafilomycin A1, a vacuolar type H⁺-ATPase inhibitor which inhibits endosomal acidification, in a dose-dependent manner (Forzan et al., 2007). The most straight-forward evidence of the conformational change according to pH change was provided in a study showing the different overall structures of whole BTV under pH

8.8 and two acidic pHs where VP5 spikes can be seen to protrude from the viral surface (X. Zhang et al., 2016). Although VP5 lacks the autocatalytic cleavage site and N-terminal myristoyl group which are the triggers during the entry proteins of reoviruses and rotaviruses, it bears two N-terminal amphipathic helices which are proved to cause leakiness of the membrane upon low-pH treatment and progressive deletion of the N terminus was found to increase the expression level by reducing cytotoxicity (S. H. Hassan, Wirblich, Forzan, & Roy, 2001). A previous study suggested VP5 associates with lipid raft domains, which are some dynamic microdomains in cellular membranes that are formed by the segregation of membrane lipids like sphingolipids and cholesterol in a glycerophospholipid-rich environment (Bhattacharya & Roy, 2008). Evidence for this includes the observation that a reduced level of cellular cholesterol leads to a decreased BTV viral titer, and that VP5 possesses a highly conserved WHXL motif which present in a SNARE (Soluble NSF Attachment protein [SNAP] Receptor) regulatory protein Synaptotagmin I (Syt1). This protein has the function of docking with the cellular protein in the plasma membrane. When a mutation of the WHXL motif into AAAA was introduced to VP5, the protein no longer localized at the plasma membrane but instead formed patches in the cytoplasm (Bhattacharya & Roy, 2008). A recent study demonstrated that when BTV penetrates the late endosomal membrane due to the late endosome-specific lipid factor 2,2'-dioleoyl lysobisphosphatidic acid (LBPA), the process is probably due to a combination of anionic charge and fluidic properties of LBPA. This study also shows that VP5 forms pores of a discrete size and that LBPA may allow these VP5 membrane pores to expand in a concentration-dependent manner (Patel, Mohl, & Roy, 2016).

The membrane penetration process of non-enveloped viruses and molecular details of membrane disruption and the structural characteristics of the domains involved vary

for different viruses but it appears to follow a general principle that the virus traffics to a penetration site and engages cellular factors that impart structural changes to enable the viral particle to bind and penetrate a target membrane.

1.4. Structural and Functional Studies of Maize Streak Virus

1.4.1. Introduction of Geminiviruses

Viruses pose serious threats to global agriculture, the yield loss imposed by viral pathogens ranks second next to pathogenic fungi, revealed by the world-wide crop productivity survey (Oerke & Dehne, 2004). Geminiviruses are plant viruses infecting a variety of crops such as cassava, cotton, legumes, maize, tomato and okra causing great economic losses and posing a serious threat to food security during the past two decades, particularly in the tropics and subtropical areas. For example, in Africa and Asia, complete losses in infected fields were caused by maize streak disease, cassava mosaic disease and cotton leaf curl disease (Shepherd et al., 2010)(Legg & Fauquet, 2004)(Sattar, Kvarnheden, Saeed, & Briddon, 2013), and the tomato yellow leaf curl disease is one of the major viral diseases of tomato worldwide (Scholthof et al., 2011).

Geminiviruses have been studied extensively at the molecular level because of their agricultural importance. They encompass seven genera: Mastrevirus, Begomovirus, Curtovirus, Becurtovirus, Eragrovirus, Topocuvirus and Turncurtovirus. The evolutionary relationships between the members of the different geminivirus genera are mapped with a phylogenetic tree (Varsani et al., 2014) (Figure 1.9). The Begomovirus can have mono- or bipartite genomes designated as DNA-A and DNA-B while the other 6 genera only have monopartite genomes(Varsani et al., 2014). Of

the 7 genera, Mastrevirus, Curtovirus and Begomovirus are the three best characterized. Mastreviruses are found primarily in the Old World, transmitted by leafhoppers and infect both monocotyledonous and dicotyledonous plants; Curtoviruses infect only dicots in the New World and are also transmitted by leafhoppers; Begomoviruses are found in the Old and New World, are transmitted by whiteflies and constitute the largest genus (Hanley-Bowdoin, Bejarano, Robertson, & Mansoor, 2013).

Geminivirus particles are characterized by a circular, single-stranded DNA genome within a twinned icosahedral capsid (W. Zhang et al., 2001). The genome sizes are from 2.5 to 5.5 Kb, indicating a highly reduced size which limits the coding capacity. However, geminiviruses have evolved overlapping open reading frames (ORFs) to compensate for these restrictions which encode between four and eight multifunctional proteins (Figure 1.10) that are able to effectively manipulate plant functions and cause devastating diseases in staple and fibre crops worldwide.

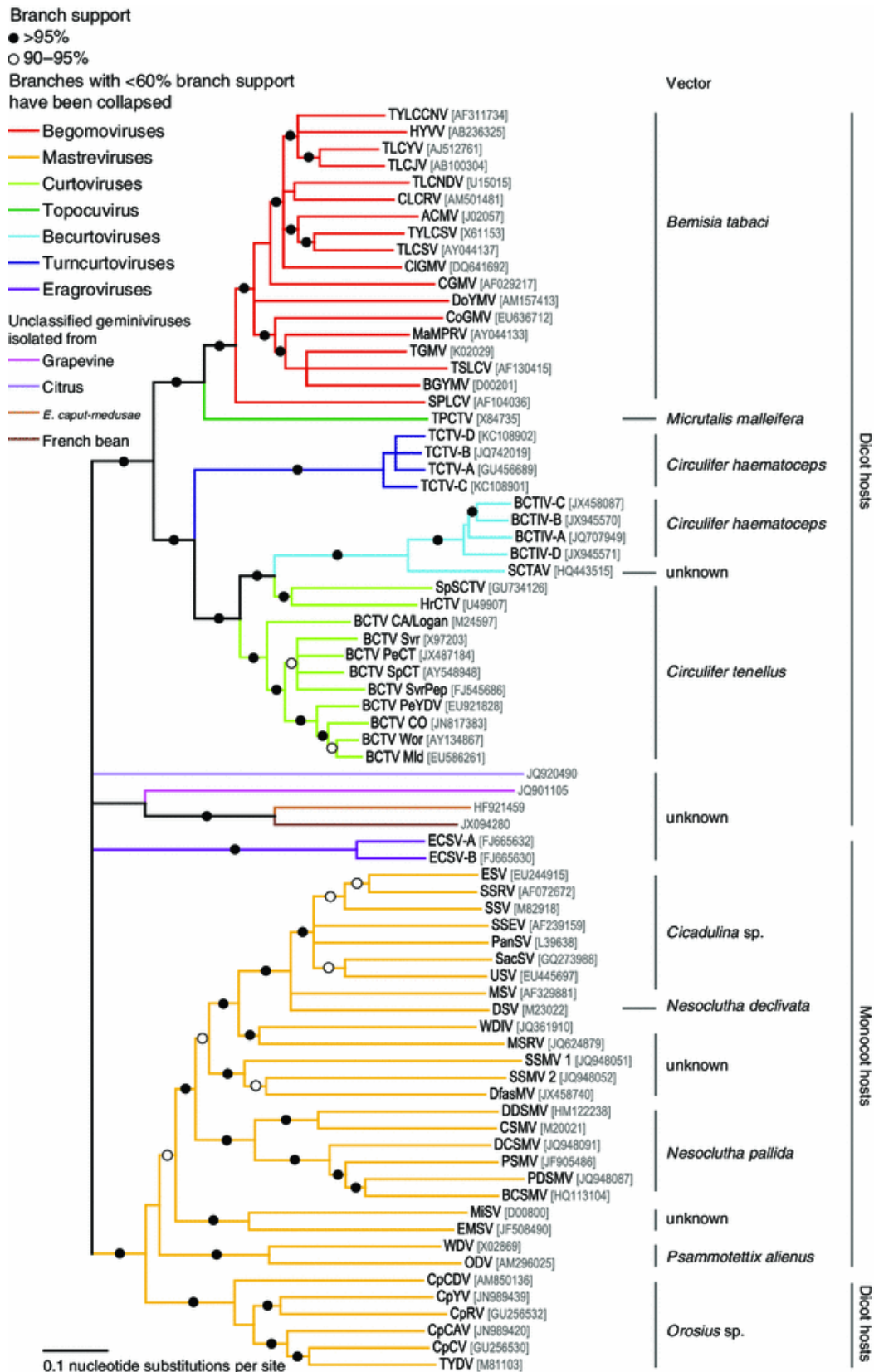


Figure 1.9 Phylogenetic tree of the full genome sequences of representative geminiviruses (Varsani et al., 2014).

1.4.2. The geminiviruses infection of plants

Geminivirus infection causes symptoms including plant stunting and a failure of reproductive organs to develop normally. Typically leaves are curled, and deformed with a yellow mosaic or mottled pattern, and sometimes vein swelling and enations (Mansoor, Briddon, Zafar, & Stanley, 2003). These changes on the plants caused by the viruses implicate extensive changes in host transcription that lead to cellular homeostasis and developmental processes being altered including thousands of differentially expressed genes involved in diverse processes from defence and programmed cell death to DNA replication and cell cycle control (Ascencio-Ibáñez et al., 2008)(Pierce & Rey, 2013). Misregulation of host miRNAs related to developmental transitions and hormone signalling is also detected during infection (Afsar R Naqvi, Haq, & Mukherjee, 2010)(Amin, Patil, Briddon, Mansoor, & Fauquet, 2011a). A schematic diagram was made to represent the role of host factors involved in the diverse cellular process during geminivirus infection which shows the cellular and metabolic pathways such as cell division and ubiquitination which are altered by viral proteins through different processes (Sahu, Sharma, Puranik, Muthamilarasan, & Prasad, 2014). The ubiquitin proteasome system is the prime defence machinery against pathogen invasion. (Citovsky, Zaltsman, Kozlovsky, Gafni, & Krichevsky, 2009)(Dielen, Badaoui, Candresse, & German-Retana, 2010). Silencing the ubiquitin activating enzyme which interacts with transcriptional activator protein (TrAP protein) of geminivirus could promote viral infection (Czosnek et al., 2013). The viruses can also take advantage of the host protein degradation processes to reduce host defense and reduce cell death, thus promoting viral replication (Citovsky et al., 2009).



Figure 1.10 Genome organisation of isolates in various geminivirus lineages. Abbreviations: LIR, long intergenic region; SIR, short intergenic region; CR, common region; rep, replication-associated protein; ren, replication enhancer; trap, transactivator protein; ss, silencing suppressor; tgs, transcriptional gene silencing; sd, symptom determinant; cp, capsid protein; mp, movement protein; reg, regulatory gene; NSP, nuclear shuttle protein (Varsani et al., 2014).

During infection, an environment suitable for virus multiplication is created by the replication initiation protein (Rep) interacting with SUMO (small ubiquitin-related modifier) conjugating enzyme (SCE) (Castillo, Kong, Hanley-Bowdoin, & Bejarano,

2004). A viral replication favoured S-phase can be initiated by geminiviruses like other DNA viruses, by binding of Rep with the viral promoter, interference with cell cycle regulatory proteins (Retinoblastoma-related; RBR/E2F pathway), and interaction with the host DNA replication factors, in which process RBR is the directing factor that controls the cell cycle regulation and differentiation (Jordan, Shen, Hanley-Bowdoin, & Robertson, 2007).

RNA silencing can be induced upon viral infection in the plant depending on the host- or virus-derived small RNAs like microRNAs(miRNAs) or small interfering RNAs(siRNAs) (Namisha, Sahu, & Puranik, 2013). On the other hand, to counteract this silencing mechanism, there are many viral RNA silencing suppressors among plant viruses which interact with key components in this process (Namisha et al., 2013). In the case of begomoviruses, a common pattern of making alterations to some key development-related genes by miRNA generation was reported during infection (Amin, Patil, Briddon, Mansoor, & Fauquet, 2011b). In tomato, predicted miRNAs which tend to bind to the ORF in the Tomato leaf curl New Delhi virus (ToLCNDV) genome and deregulate the disease-related mir159/319 and mir172 genes were reported (Afsar Raza Naqvi, Choudhury, Mukherjee, & Haq, 2011). The exact functions of a large number of plant miRNAs remained unidentified.

1.4.3. Genimivirus interaction with the insect hosts

Compared to the alteration done with plants hosts, the interaction with the insect hosts is less studied for geminiviruses but some recent research has shown there are some changes in signalling and defence pathways because of the infection. The studies of virus-insect interaction are mainly on begomoviruses and their *Bemisia tabaci* (whitefly) vector, which acquires viruses during feeding on the phloem of an infected

plant. More than 1600 genes differ in expression between virus-carrying and non-carrying whiteflies (Luan et al., 2011) which represent diverse pathways such as cell cycle, protein synthesis and lipid metabolism. More precisely for genes involved in the immune response, the presence of virus can activate all of the autophagy genes and most genes associated with lysosome function while downregulation occurs for genes involved in apoptosis and signal transduction of the immune response.

Because of the very small genome sizes of geminiviruses, they have only four to eight proteins. To make the viruses more genetically economical, there are mainly two strategies taken: one is gene overlapping, the other is protein multifunctionality by bringing together functional domains into a single polypeptide chain, such as replication initiation protein (Rep/RepA), transactivator protein (TrAP), and coat protein (CP).

1.4.4. Introduction of the capsid protein (CP) of geminiviruses

The CP is the only structural protein of geminivirus particles (Fondong, 2013). The CP protein is multifunctional, it not only packages viral genome inside particles, but is also involved in cellular spread of virus, shuttling of viral DNA in and out of the nucleus, (H. Liu, Lucy, Davies, & Boulton, 2001), and insect transmission. The latter was tested with the whitefly-transmitted ACMV and the leafhopper-transmitted BCTV which showed that an ACMV chimeric virus containing the BCTV CP gene is transmitted by leafhoppers which naturally transmit BCTV but not ACMV (Bridson, Pinner, Stanley, & Markham, 1990). Studies in mastreviruses and Tomato yellow leaf curl virus (TYLCV) indicated that some geminivirus CPs transfer virus DNA between the nucleus and the cytoplasm (H. Liu et al., 2001; Rojas et al., 2001). The N-terminal domain of the CP which binds to ssDNA is responsible for the nuclear import and

export. In the tomato yellow leaf curl Thailand virus this is the N-terminal 62 amino acid residues (Pitaksutheepong, Vimoltat, Nathwong, & Attathom, 2007). In MSV, the CP protein has been shown to bind, in a sequence non-specific manner, both ss and ds viral and plasmid DNA by a binding domain determined within the 104 N-terminal amino acids (H Liu, Boulton, & Davies, 1997). This provides evidence that the CP is responsible for viral DNA nuclear transport at the same time as encapsidation, which plays a role in intra- and inter-cellular movement in systemic infection. Protein sequences of Panicum streak virus (Mastrevirus), Mesta yellow vein mosaic virus (Begomovirus), Horseradish curly top virus (Curtovirus) and Tomato pseudo-curly top virus (Topocuvirus) were structurally predicted by the MetaServer (Ginalski, Elofsson, & Fischer, 2003). The icosahedral virus structures CP of Satellite tobacco necrosis virus (STNV) was found by the software to be the only suitable template for structural modeling with significant scores for all four geminiviral CPs. In all these structures predicted, an eight stranded (β B- β I) β -barrel fold (with two sheets BIDG and CHEF), which is also known as the viral jelly-roll can be seen (Krupovic, Ravantti, & Bamford, 2009) (Figure 1.11).

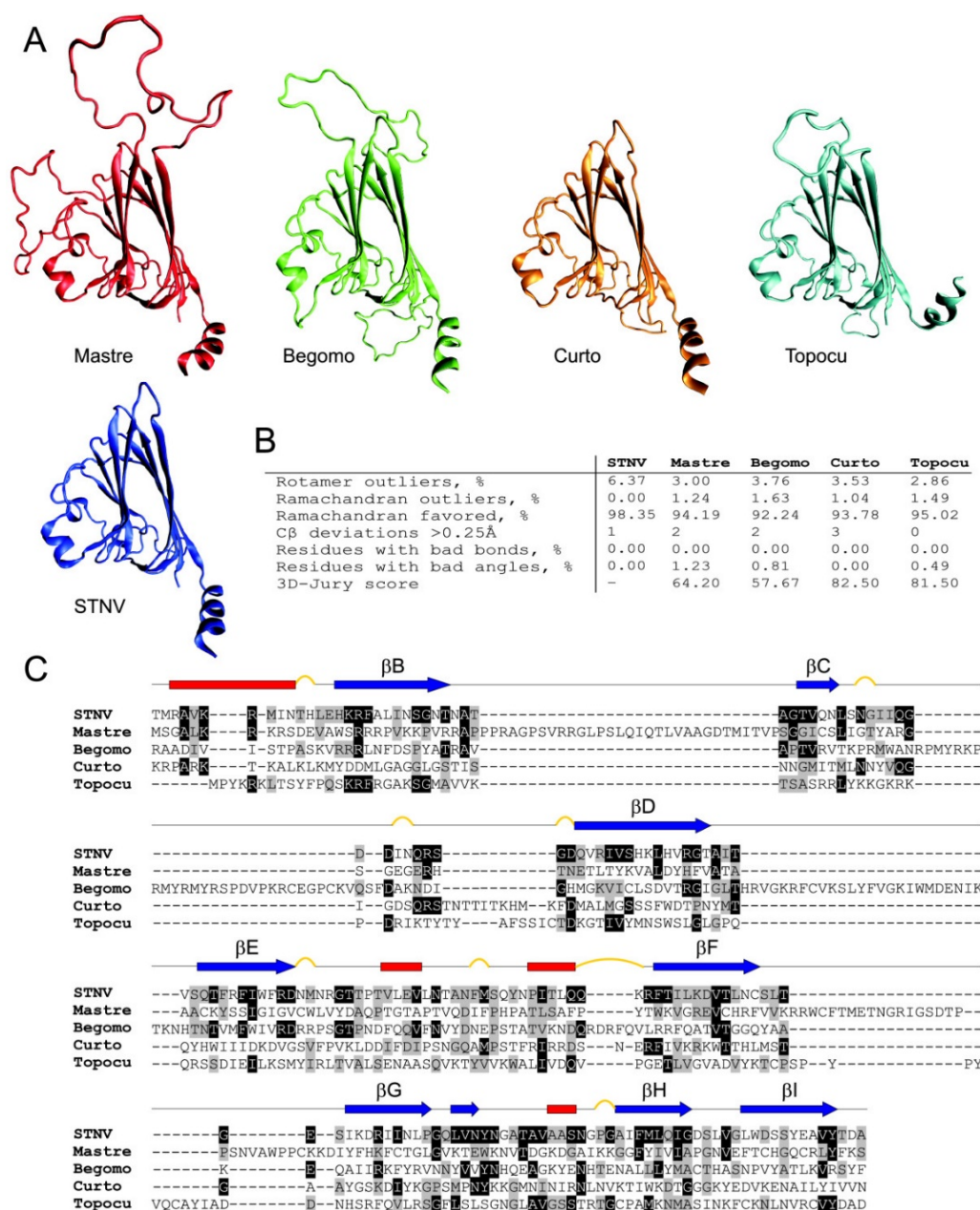


Figure 1.11 Structural modeling of the geminiviral capsid proteins (CPs). (A) Pseudo-atomic models of the CPs from genera of Mastre (*Panicum streak virus*), Begomo (*Mesta yellow vein mosaic virus*), Curto (*Horseradish curly top virus*) and Topocu (*Tomato pseudo-curly top virus*) compared to the atomic model of STNV CP. (B) Comparison of the stereochemical quality of the STNV CP X-ray structure to that of the pseudo-atomic models of geminiviral CPs. (C) Alignment of the CPs comparing to CP of STNV (Krupovic et al., 2009).

1.4.5. Introduction of their proteins in geminiviruses

A replication-associated protein (Rep) is expressed by all geminiviruses from ORFs C1/C2 (Figure 1.10) (Varsani et al., 2014) which is essential for replication. Except for mastreviruses and becurtoviruses, all other viruses in the family express their Rep from a single intronless C1 ORF. After infecting cells, the single-stranded (ss) genomes of geminiviruses are amplified with a combination of rolling-circle and recombination-mediated replication (Gutierrez, 1999). The conserved protein Rep mediates initiation of viral replication and depends on host replication machinery to copy their genomes rather than acting as a DNA polymerase itself (Heyraud-nitschke et al., 1995). Among the different genera of geminiviruses, although mastreviruses have an even smaller inventory of genes than other geminiviruses, they still utilize multiple pathways of replication as other genera such as begomoviruses and curtoviruses do. The replication products including ssDNA, open circular DNA (ocDNA), covalently closed circular DNA (cccDNA) and linear heterogeneous-length dsDNA (hdsDNA) molecules in both monomeric and multimeric states as well as several true replicative DNA intermediates of complementary-strand replication (CSR), rolling-circle replication (RCR) and recombination- dependent replication (RDR) have been seen (Erdmann et al., 2010) (Figure 1.12).

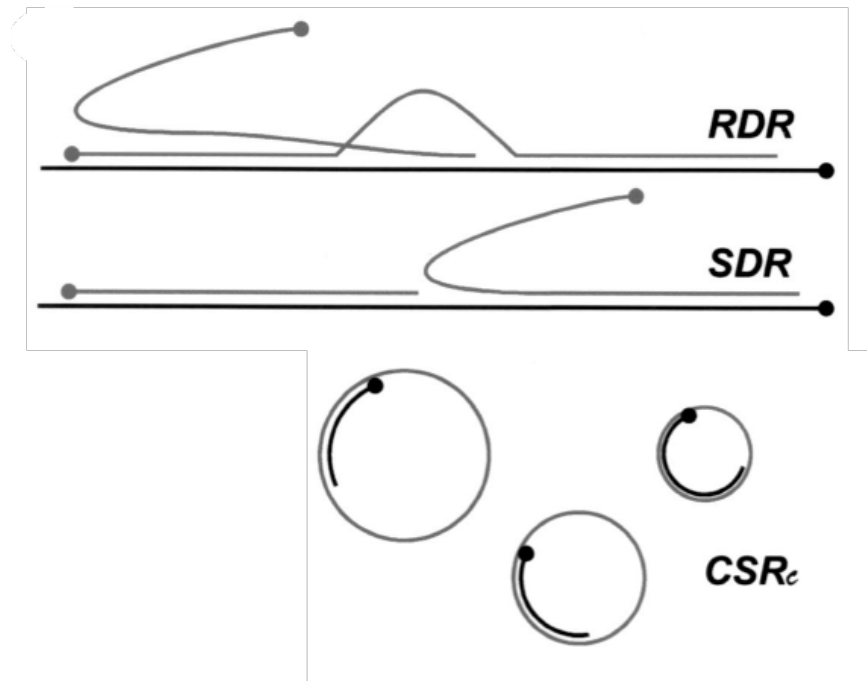


Figure 1.12 Schematic summary of viral DNA replication forms. (Preiss & Jeske, 2003).

The Rep is multifunctional, it can mediate virus-specific recognition of its cognate origin of replication (Castillo et al., 2004), facilitate initiation and termination of viral DNA synthesis (Laufs, Traut, et al., 1995), and accumulates host replication factors in viral-infected plant cells (Nagar, Pedersen, Carrick, Hanley-Bowdoin, & Robertson, 1995). Rep can recognise a specific repeating sequence at the 5' intergenic region in the dsDNA and bind to it (Singh, Malik, Choudhury, & Mukherjee, 2008), followed by cleaving and ligating DNA at a hairpin loop of the plus-strand origin and act as a helicase to unwind the DNA while replicating the genome (Cl  rot & Bernardi, 2006; Singh et al., 2008). Rep also interacts with a number of proteins. (Choudhury et al., 2006; Orozco, Kong, Batts, Elledge, & Hanley-Bowdoin, 2000). It interacts with AL3 to achieve high level viral DNA accumulation (Elmer et al., 1988), with coat protein to interfere with DNA cleavage and ligation activities in Mung bean yellow mosaic India virus (Malik, Kumar, Bagewadi, & Mukherjee, 2005). Rep also interacts with

several host factors like clamp proliferating cell nuclear antigen (PCNA), the clamp loader replication factor C (RFC), the ssDNA binding protein RPA, histone H3, and a mitotic kinesin (Bagewadi, Chen, Lal, Choudhury, & Mukherjee, 2004; Kong & Hanley-Bowdoin, 2002; Malik et al., 2005; Singh, Islam, Choudhury, Karjee, & Mukherjee, 2007), all of which are factors involved in DNA transactions, with host proteins for cell cycle regulations like retinoblastoma protein (RBR) (Grafi et al., 1996).

The movement protein (MP) is required for virus cell-to-cell movement (Ward, Medville, Lazarowitz, & Turgeon, 1997). It is encoded by the V2 ORF in monopartite begomoviruses and many other genera. MP interacts with plasmodesmata to realize inter-cellular transport (Rojas et al., 2001). MP functions of monopartite begomovirus and of mastrevirus are different. Monopartite begomovirus MPs bind viral ssDNA like bipartite begomovirus MPs, while mastrevirus MPs do not bind viral DNA, instead, the transferring of DNA is facilitated by interact with the CP protein which binds to DNA and forms a movement complex (H. Liu et al., 2001). In Tomato leaf curl Java virus (ToLCJV), a nuclear export signal (NES) was mapped to the N-terminus of MP which suggests it has an enhancement function in CP-mediated viral DNA nuclear export (Sharma & Ikegami, 2010). In TYLCV, MP forms nucleoprotein complexes together with CP, MP and C4, to deliver viral DNA to the plant cell periphery (Rojas et al., 2001). The ER system is important for MP intracellular trafficking in several geminiviruses. As an example in SqLCV (a phloem-limited virus) and CaLCV (which infects all cell types), the MPs bind to ER-derived tubules in undifferentiated vascular tissues (procambium) which penetrate the cell wall (Ward et al., 1997), while in Abutilon mosaic virus (AbMV) (a phloem-limited virus), the MP has been shown to travel from the ER to the plasma membrane using the cellular membrane flow (S. C.

Zhang, Ghosh, & Jeske, 2002). In the case of the Euphorbia mosaic virus (Kim & Lee, 1992) and Indian cassava mosaic virus (Roberts, 1989), these viruses use microtubules to transport MP complexes between cells instead of the ER system.

Apart from the Rep, CP and MP proteins described above which are expressed in all genera of geminiviruses, there are some proteins that are only produced in some particular viruses. The mastrevirus has the smallest genome which only expresses four proteins. Rep and RepA protein encoded by the C1 and C2 ORFs on the complementary (C) sense transcripts are for viral replication; The V1 and V2 ORFs virion (V) sense transcripts encode CP and MP proteins which are required for systemic infection and subsequent disease development (Fondong, 2013), separated by two intergenic regions (IRs) on opposite sides of the genome which are large (LIR) and small (SIR) (Figure 1.12) (Kammann et al., 1991).

Transcriptional activator protein (TrAP) is produced by several geminiviruses, including monoparticle begomovirus (C2 ORF), bipartite begomovirus (AC2 ORF), curtoviruses (C2 protein) and topocuviruses (L2 protein) (Fondong, 2013). TrAP is best studied in begomoviruses as a multifunctional protein that participates in virus pathogenicity (Hong, Saunders, Hartley, & Stanley, 1996), gene activation (Shivaprasad et al., 2005), and gene silencing suppression (Chowda-Reddy et al., 2009). The transcription of CP and MP can be activated by a multimer formed by TrAP when expressed in insect cells. Upon infection, TrAP is phosphorylated and contains an C-terminal acidic transcriptional activation domain and a zinc-finger. The minimal activation domain is quite small, the 15 C-terminal amino acids of the protein are alone capable of transcription activation in mouse fibroblasts (NIH3T3 cells) when fused to a heterologous DNA-binding domain (Hartitz, Sunter, & Bisaro, 1999).

A replication enhancer protein (REn) is encoded in bipartite begomoviruses (ORF AC3) and in curtoviruses and monopartite begomoviruses (ORF C3) which enhances viral DNA accumulation and symptom development in infected plants by interacting with Rep and PCNA (Stanley & Latham, 1992; Sung & Coutts, 1995). This role of REn was further confirmed in the Tomato leaf curl Kerala virus which showed an increased activity of Rep-mediated ATPase by the interaction of REn with Rep (Pasumarthy, Choudhury, & Mukherjee, 2010).

Except for mastreviruses which lacks a C4 ORF, the C4 ORF in monopartite geminiviruses and AC4 in bipartite geminiviruses encode proteins that are the least conserved among geminiviruses. A viral DNA level reduction and disruption of symptom development can be caused upon mutation of the C4 ORF in BSCTV (Teng et al., 2010), TYLCV (Jupin, De Kouchkovsky, Jouanneau, & Gronenborn, 1994) which further implied the protein is related to viral movement and infection symptom development. C4 can also serve as an indicator of vein swelling phenotype in BCTV infected plants and the expression of the protein in transgenic *N. benthamianacan* can cause ectopic cell division which may be the reason for the vein swelling (Mills-Lujan & Deom, 2010; Piroux, Saunders, Page, & Stanley, 2007). The expression of RKP, ATHB7 and ATHB12 induced by BSCTV C4 in infected plants can be related to regulation of cell cycle, cell development and differentiation (Lai et al., 2009; Park et al., 2011).

In bipartite begomoviruses, the function of protein expressed by ORF AV2 is not well defined while the corresponding V2 ORF in TYLCV of the monopartite begomovirus and BCTV of the curtovirus are possible pathogenicity determinants (Hormuzdi & Bisaro, 1995; Wartig et al., 1997). The protein expresses by the V2 ORF in monopartite begomovirus and mastrevirus is the movement protein (MP) described

above. The plants infected with V2 or AV2 mutated in Papaya leaf curl virus or Cotton leaf curl Kokhran virus had lower levels of viral DNA in plants (Mubin, Amin, Amrao, Briddon, & Mansoor, 2010), the same results were also shown in ToLCV and ToLCNDV (Padidam, Beachy, & Fauquet, 1996; Selth, Randles, & Rezaian, 2004).

In the DNA-B of bipartite begomoviruses, a BV1 ORF encodes the BV1 protein which is a nuclear shuttle protein (NSP) that traffics viral ssDNA between the nucleus and the cytoplasm, coordinate the movement of viral DNA across both nuclear and plasmodesmal boundaries (Noueiry, Lucas, & Gilbertson, 1994) by forming a DNA–NSP complex in a sequence non-specific manner or interact with MP to achieve intracellular movement across the adjacent cell walls and then the complex is directed to the nucleus to initiate replication in the new cells (Lazarowitz & Beachy, 1999). In CaLCV, the NSP works by interacting with a nuclear shuttle protein interactor Arabidopsis protein to shuttle viral DNA between the cytoplasm and cell nucleus (McGarry et al., 2003). In some geminiviruses, NSP also interacts with NIG (NSP-interacting GTPase) to promotes the translocation of the viral protein from the nucleus to the cytoplasm and its redirection to the cell surface to interact with the viral movement protein, MP (Carvalho, Machado, Zerbini, & Fontes, 2008).

1.4.6. Introduction to Maize Streak Viruses

Maize streak virus (MSV) is the type member of genus mastrevirus of the family Geminiviridae with a monopartite genome of single-stranded circular DNA. MSV infects maize and causes the devastating maize streak disease (MSD) throughout sub-Saharan Africa. It also infects more than eighty other plant species like wild grasses in the family Poaceae (Harkins et al., 2009). Of the eleven known MSV strains only MSV-A infects maize and causes serious disease (Martin, Willment, & Rybicki, 1999;

Monjane, van der Walt, Varsani, Rybicki, & Martin, 2011). In 1901, Claude Fuller recorded MSD for the first time in South Africa motivated by an on-going outbreak (Shepherd et al., 2010). MSD was shown to be caused by a virus and transmitted by a leafhopper by Storey (Storey, 1924) and proved by visualization of the novel, twinned, quasi-icosahedral (geminata) shaped virus particles in maize tissue in 1974 and the name of 'geminivirus' was given (Bock, Guthrie, & Woods, 1974). This was the first time a circular single-stranded DNA had been found in plant viruses (Goodman, 1977; B. D. Harrison et al., 1977). The leafhoppers initially transmit the geminiviruses particles into the phloem sieve tubes and the viruses are then concentrated in the mesophyll cells of infected leaves which can cause malformation of chloroplasts and reduced chlorophyll production that forms the pattern of streaks (chlorotic lesions) in the infected leaves (Lucy, Boulton, Davies, & Maule, 1996) and the leafhoppers acquire viruses from the lesions (Peterschmitt, Quiot, Reynaud, & Baudin, 1992).

MSV is transmitted by different species of *Cicadulina* with different proportions from 15% to 45% of individuals capable of transmitting the virus (Mesfin, Den Hollander, & Markham, 1995) and only some of the 18 species of *Cicadulina* can transmit MSV (Lett et al., 2002). Among the non-vector species, the gut wall of the insects is the barrier to transmission confirmed with PCR detection showing that MSV was restricted to the gut of non-vector species of *C. chinai* but can be detected throughout the gut, haemolymph and head of *C. mbila*, a vector species, less than 3-hours' post injection that indicates a transmembrane mechanism through a receptor (Lett et al., 2002). Although the viral level can be kept for some time, the virus MSV is thought not to replicate in leafhopper vectors (Peterschmitt et al., 1992). To reduce the spread of by leafhoppers movement, bare ground barriers can be set between early- and late-planted maize fields (Bosque-perez, 2000). Rotation planting of crops and avoiding

planting new maize at the downwind side of the old batch can also limit the movement of leafhopper vectors (Rose, 1978). Seed treatment with chemicals and usage of pesticides before or during planting also provide options for farmers. The most effective way to prevent MSD is probably by researching and application of streak-resistant cultivars which can reduce symptom severity, lower virus titres, consequently reduce virus incidence in the field.

MSV has a single-stranded (ss) DNA genome of 2.7 kb containing four open reading frames, a long intergenic region (LIR) and a short intergenic region (SIR). The four ORFs encodes four proteins, a coat protein (CP) from the V1 ORF, a movement protein (MP) from the V2 ORF, and two replication-initiator proteins of Rep encoded from the C1 and C2 ORFs and RepA from C1 ORF alone. CP and MP are from the virion-sense strand while the Rep and RepA are from the complementary strand. They share the same N-terminus but differ at the in C-termini which results in their different functions in the virus. The LIR contains a hairpin structure that can be cleaved by Rep to initiate rolling circle replication (Laufs, Jupin, et al., 1995). A new circular ssDNA progeny is generated and can be converted to dsDNA by complementary-strand replication (CSR) and then into cccDNA when it is packaged around host nucleosomes (Aktories et al., 2010; Saunders, Lucy, & Stanley, 1991). RepA is involved in induction of the virus replication environment in infected cells but the specific role needs further study (Gutierrez, 2000).

Begomovirus and curtovirus use three replication modes of replication (complementary-strand replication [CSR], rolling-circle replication [RCR] and recombination-dependent replication [RDR]), and MSV has been shown to do the same, although mastreviruses have fewer regulatory genes than other geminiviruses

and the replicative intermediates of CSR, RCR and RDR were detected within infected maize tissues (Erdmann et al., 2010).

1.4.7. Introduction of the CP of MSV and the structure of the viral capsid

CP of MSV is a basic protein with a pI of 10.8 that binds to both ss and ds viral and plasmid DNA in a sequence non-specific manner (H Liu et al., 1997), which indicates that in addition to the role of genome encapsidation, the CP may also play a role in viral DNA transport. The maize infectivity of MSV was abolished when the CP gene was mutated or deleted and no accumulation of ssDNA occurred in protoplasts anymore (M. I. Boulton, Pallaghy, Chatan, Macfarlane, & Davies, 1993; M. Boulton et al., 1989). The 104 N-terminal amino acids of the protein were mapped to be the DNA binding domain although no consensus DNA binding motif was determined by computer-assisted analysis (Varagona & Raikhel, 1994) but 9 basic amino acids are contained within the residues 5–22 which may be responsible for DNA binding. In MSV, the CP also binds to MP as described in the previous paragraphs which is required for systemic infection. The MP does not bind to viral DNA as CP does but it forms an MP–CP complex and thus prevents accumulation of CP–DNA complexes in the nucleus of microinjected plant cells (Huanting Liu, Oparka, & Davies, 2001). A model was proposed that the MP-CP complex blocks the signal by which CP transfers viral DNA into the nucleus, instead redirecting the DNA to the cell periphery for movement to neighbouring cells (Huanting Liu et al., 2001).

In 2001, a 25 Å resolution structure by cryo-electron microscopy (cryo-EM) and three-dimensional image reconstruction methods of frozen-hydrated MSV-N particles was reported (W. Zhang et al., 2001). The MSV capsid consists of two incomplete icosahedra of dimensions 220 x 380 Å with fivefold symmetry axis coincident with

the long axis of the particle, so that the CP are arranged with quasi-icosahedral symmetry (W. Zhang et al., 2001). A total of 110 copies of CP are arranged with 52-point symmetry, with each half of the particle is composed of 11 pentameric capsomers. During reconstruction, fivefold symmetry was imposed so the apical capsomers which are on the two ends of a MSV particle were reconstructed with strict five-fold symmetry. The 10 peri-pentonal capsomers are the ones next to the apical capsomers and they appeared to maintain intrinsic fivefold symmetry while the 10 capsomers near the equatorial region of the particle were observed to be distorted (

Figure 1.13). The gap between the two halves of the particle measured 25 Å (W. Zhang et al., 2001). In 2004, Robbie J. Reutzel in his thesis described two unpublished structures of the MSV geminate particle at an improved resolution of 9.3 Å and the MSV isometric particle at 9.7 Å (Figure 1.14), with a comparison of the two structures from which he proposed a possible conformational change of the N-terminal helix of the CP that may facilitate the assembly mechanism. In 2013, Kyle Dent from University of Cape Town in his thesis performed a 3D reconstruction of both the geminate and isometric MSV using standard single-particle image processing methodology, with resolutions of 26 Å of the geminate from a negative-stain dataset, 23 Å from a cryo-EM dataset, and about 15 Å for the monopartite reconstruction. Comparative modelling of the MSV CP was undertaken using the pentamer (CPs) of Satellite tobacco necrosis virus (STNV) as a structural template.

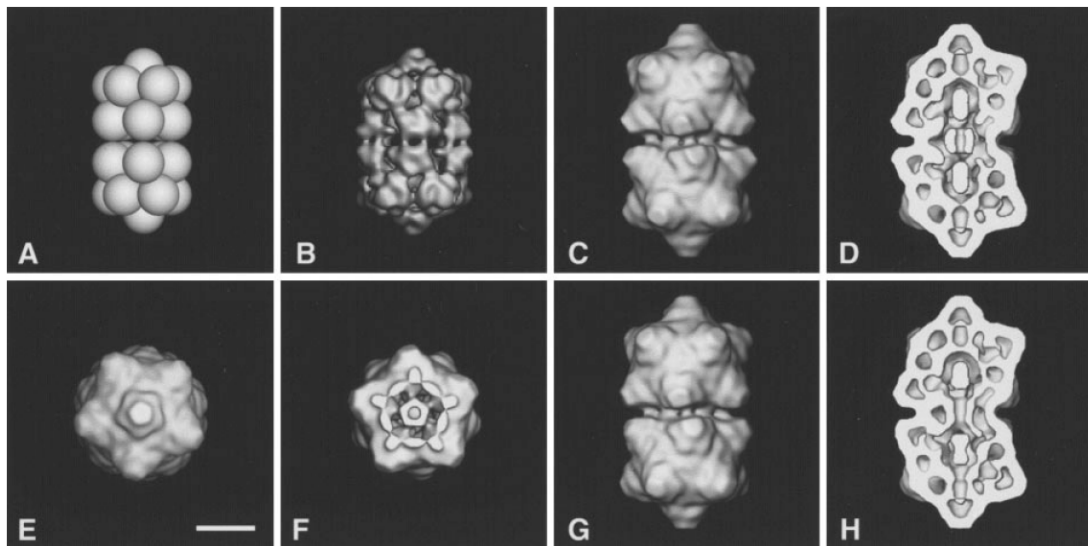


Figure 1.13 Models and image reconstruction of MSV-N. (A) “Ping-pong” model for the capsid structure of gemini viruses, consisting of 22 “subunits” (capsomers), arranged with 52-point-group symmetry. (B) Surface-shaded, pseudo-atomic model of MSV-N based on the structure of STNV rendered at 25-Å resolution and used as an initial model for particle orientation refinement. (C) MSV-N reconstruction computed with 52-dihedral-point-group symmetry and viewed along a twofold axis of symmetry. (D) Same as (C) but with the nearest half of the reconstruction removed to expose the particle interior. (E) End-on-view of MSV-N reconstruction in which only fivefold axial symmetry was imposed during the processing. (F) Same as (E) but with one head removed to reveal features in the gap region. (G, H) Views of the reconstruction with just fivefold cyclic symmetry for comparison with corresponding views of the 52-symmetric reconstruction shown in (C) and (D) (W. Zhang et al., 2001).

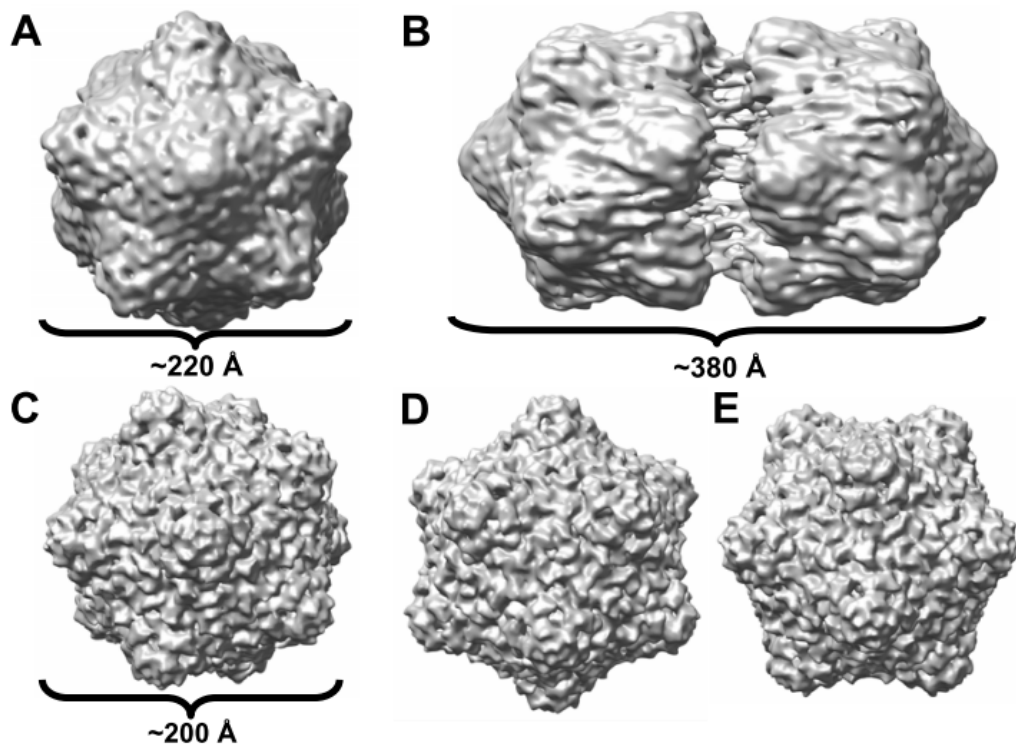


Figure 1.14 Views of Geminate and Single MSV Capsids Down Symmetry Axes. (A) 5-fold view of geminate capsid (B) 2-fold view of geminate capsid (C) Icosahedral 5-fold view of single capsid (D) Icosahedral 3-fold view of isometric particle (E) Icosahedral 2-fold view of T=1 isometric particle (Reutzel thesis, 2005).

Chapter 2.

Materials and Methods

2.1. Constructs Design

2.1.1. Protein sequence alignment and phylogenetic analysis of 26 serotypes of BTV

The VP5 proteins of the 26 serotypes of BTV were aligned using the online program ClustalW2 (EMBL-EBI) and displayed with the online program ESPript (esprict.ibcp.fr/) (Robert & Gouet 2014). The alignment was used to construct a phylogenetic tree and the numbers shown on the branches marked the distance in sequence similarity. The grouping is the guide for us to choose the serotypes from which VP5 constructs were designed.

2.1.2. Constructs design

The pOPIN vector system created by the Oxford Protein Production Facility (OPPF) was used for ligation-independent cloning (LIC) (Berrow et al. 2007). The pOPIN vectors utilize multiple promoter systems to allow protein expression in different cell types and a *lacZ* insert for blue/white screening during cloning. In summary, both N- and C- terminal hexa-his tags were added to full-length VP5, and two N- terminal truncations of 21 and 41/43 amino acids. These N-terminal truncations are designed to remove either 1 or 2 amphipathic α -helices which have been shown to have a deleterious effect on protein expression. The boundaries of the constructs were defined using helical wheel projections (Figure 2.1) and previous results (Hassan et al. 2001). A series of constructs of VP5 proteins from different BTV serotypes were designed. The DNA for the constructs was amplified by PCR from the synthesized full-length gene (GeneArt). Genes for different constructs were then cloned into the appropriate pOPIN vectors using the In-Fusion system.

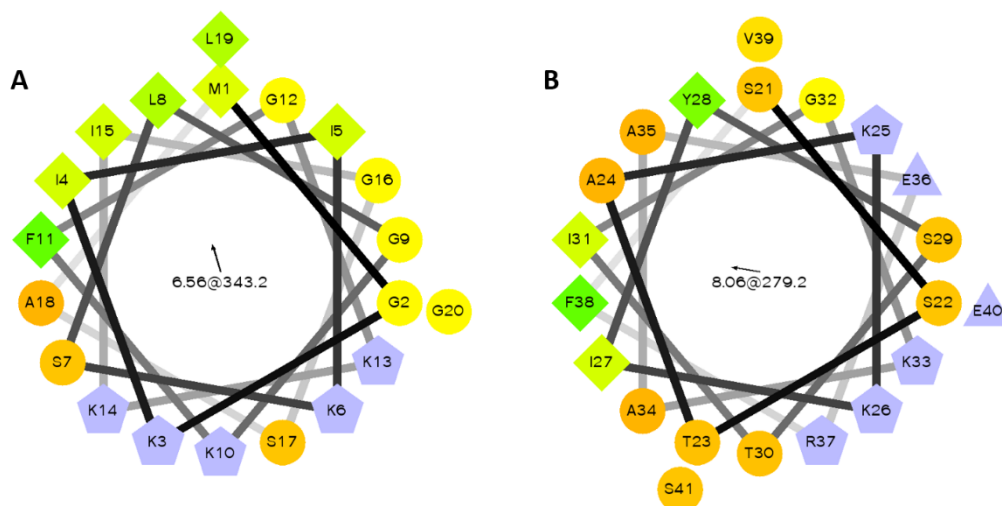


Figure 2.1 Helical wheel representations of (A) aa 1 to 20 and (B) aa 21 to 41 of BTBV15 VP5. The program Helical Wheel was used to plot a helical wheel representation of the N-terminal amino acids of VP5. Each panel represents a proposed α -helix viewed along the helix axis with the indicated amino acid residues. Hydrophobicity is shown as shades of green and the hydrophobicity decreases proportionally to the amount of green, yellow represents zero hydrophobicity. Hydrophilic residues are depicted as shades of red, pure red most hydrophilic. Light blue represents potentially charged residues.

2.2. High-throughput cloning

Productivity and efficiency of protein production have been markedly increased by the parallel development of pipelines and workflows for protein genomics. The methods used here for cloning largely utilized the automated protocols from the Oxford Protein Production Facility (OPPF). They are part of the standard protocols described in the OPPF pipeline procedure (Berrow et al. 2007). The vectors provided by OPPF are characterized as ‘pOPIN vectors’, which can be used with the commercially available InFusionTM (Clontech Laboratories, Inc.) ligation independent cloning (LIC)

techniques. A list of pOPIN vectors can be found online at <https://www.oppf.rc-harwell.ac.uk/OPPF/protocols/cloning.jsp>. The two pOPIN vectors used here (pOPINE and pOPINF) were derived from pTriEx2 backbone with an ampicillin resistance (Amp^R) and a hexa-Histidine-tag (His₆) on the C- and N-terminal, respectively (Berrow et al. 2007).

2.2.1. Primer design

Online primer design software OPINE was used for designing primers compatible with pOPINE and pOPINF vectors. The forward and reverse primers complementary to template DNA sequences of the protein constructs were designed with a standard melting temperature of around 60°C and GC content of approximately 50%. Lyophilised oligonucleotides were synthesized by MWG Eurofins and dissolved to a final concentration of 10 µM and stored at -20°C for future use.

2.2.2. DNA amplification by polymerase chain reaction (PCR)

KOD Hot Start DNA Polymerase (Novagen) was used to amplify the template DNA for achieving high fidelity. The reactions were prepared on ice in PCR tubes which were composed from the mixture listed in Table 2.2. Approximately 100 ng of template DNA was added into each PCR reaction and 34 thermal cycles were used to amplify the PCR product following the procedure listed in

Table 2.2. The PCR products were then verified by agarose gel electrophoresis and gel purified with the QIAquick gel extraction kit (Qiagen) to remove impurities.

Table 2.1 PCR reactions.

Component	Volume	Final
Total Volume 50 μL		
10x Buffer for KOD Hot Start DNA Polymerase	5 μ L	1x
25 mM MgSO₄	3 μ L	1.5 mM
dNTPs (2 mM each)	5 μ L	0.2 mM
PCR Grade Water	32 μ L	
Sense (5') Primer (10 μM)	1.5 μ L	0.3 μ M
Antisense (3') Primer (10 μM)	1.5 μ L	0.3 μ M
KOD Hot Start DNA Polymerase (1.0 U/μL)	1.0 μ L	0.02 U/ μ L
Template DNA	1.0 μ L	

Table 2.2. Thermal cycler procedure for amplifying template DNA

1. 95°C	2 minutes
2. 95°C	30 seconds
3. 60°C	30 seconds
4. 70°C	90 seconds
5. Go to step 2	Repeat for 34 cycles
6. 72°C	10 minutes
7. 4°C	Hold to end

2.2.3. High-throughput cloning with InFusion™ system

The InFusion™ system used by the OPPF pipeline is capable of incorporating target genes into vectors via homologous recombination. 100 ng of linearized pOPIN vectors and 10 ng of PCR product were mixed into lyophilized InFusion™ enzyme with a total volume of 10 µL followed by pipette mixing. The reaction is proceeded by incubation at 42°C for 30 minutes and terminated by adding 40 µL TE. The mixture can be used for transformation into competent *E. coli* cells or kept at -20°C for future use.

2.2.4. Transformation of plasmid/ InFusion™ mixture into competent cells and PCR verification

The transformation of InFusion™ reaction into *E. coli* competent cells followed the protocol of OPPF. Briefly, 5 µL of InFusion™ reaction mixture were added into 30 µL aliquots of competent cells and incubated on ice for 30 minutes. The cells were then heat-shocked for 40 seconds at 42°C by placing the whole rack in the 42°C water bath. The cells were returned to ice for 2 minutes. 300 µL of Power Broth (PB)/ Super Optimal Broth (SOC) (with no antibiotics) were added to each tube of transformed cells. The cells were transferred to a 37°C shaking incubator and incubated for 1 hour. Whilst incubating, LB Agar plates were prepared. A volume of 10ml of the agar/antibiotic/X-gal/IPTG mix was needed per plate. The stocks were diluted as follows: Carbenicillin (Cb) 50 mg/ml, 1:1000, or Kanamycin (Kan) 35 mg/ml, 1:1000, 20% X-Gal [in DMF], 1:1000, IPTG 1M stock; 1:1000 in warm agar before pouring. Plates were poured and were allowed to cool down and solidify before 50µL of cells were applied to each of the agar plates and spread evenly with spreaders. At least 10-15 minutes were allowed for the plates to dry before turning. The prepared plates were incubated overnight at 37°C in a static incubator. Blue colonies should constitute much less than 10% if the reactions were successful. The blue colonies are derived from

inefficiently linearised parental plasmid and are non- “recombinant”. Two white colonies per construct were picked from the plates which will normally give positive clones. Positive clones were confirmed by PCR and sequencing.

2.3. Protein expression in *E. coli*

For protein expression, an isopropyl- β -D-thiogalactopyranoside (IPTG)-induced *E. coli* system was used with BL21 (DE3) cells, whereby a *lac*-controlled operon that encodes bacteriophage T7 RNA polymerase contained in the host’s genome is expressed in a galactose or IPTG-dependent manner. A single colony, picked from a freshly streaked plate of the expression host carrying the recombinant vector, was inoculated into 20 mL of lysogenic broth (LB) supplemented with 50 μ g/mL carbenicillin for 4-8 hours at 37°C as a starter culture. The starter culture was expanded to 200 mL until the culture density reached an OD₆₀₀ of 0.5-0.6 and for every 1L of expression culture, a 10mL expanded culture was added and the culture was incubated with shaking at 120 rpm until the OD₆₀₀ reached 0.6. The culture was then cooled down to 4°C by placing the flasks in a cold room and IPTG was added to the culture to a final concentration of 1 mM. The induced culture was further incubated at 16°C with a lower shaking speed of 120 rpm overnight (~10 hours).

2.4. SDS-PAGE and western blot analysis

2.4.1. SDS-PAGE analysis

To prepare samples for SDS-PAGE analysis, 9 μ L aliquots of samples were mixed with 3 μ L loading buffer (10% w/v SDS; 10 mM beta-mercapto-ethanol; 20 % v/v Glycerol; 0.2 M Tris-HCl, pH 6.8; 0.05% w/v Bromophenolblue). The mixtures were

heated to 99°C for 5-10 min before being applied to the SDS-PAGE gel prepared with 5% Bis-Tris stacking gel and 12% separating gel. BenchMark™ Protein Ladder (Thermo Fisher Scientific) standards were run in parallel for protein molecular weight reference.

2.4.2. Western blot analysis

For western blot analysis, samples were collected and applied to a 4-12% SDS-PAGE gradient gel. Proteins on the gel were transferred for 85 minutes onto a nitrocellulose membrane using a Trans-Blot SD semi-dry transfer cell (Biorad) apparatus using a buffer containing 25 mM Tris-base, 350 mM MeOH and 200 mM glycine. Prestained protein marker (Invitrogen) was used to visualize successful transfer of protein to membrane. The membrane was first blocked with 5% milk in phosphate-buffered saline (PBS) buffer (12 hours, 4 °C), and gently agitated. Following blocking, the membrane was probed with mouse anti-pentaHistidine antibody (PentaHis, Qiagen) diluted 1/1,000 in blocking solution (1 hour, 25 °C). To remove non-specifically bound primary antibody, the membrane was washed with PBS buffer supplemented with 0.1% Tween-20 (PBST, 3 x 10 min). The primary antibody was then probed with IgG Fc specific anti-mouse antibody (Sigma) (1:10,000 dilution in 5% PBS-milk) for 1 hour at 25 °C. The non-specifically bound secondary antibody was washed with PBST buffer (3x10 minutes). Membrane was then soaked in 4 ml ECL reagent mixture (Amersham) and protein expression was visualized by exposure to film (Hyperfilm ECL, Amersham).

2.5. MSV Capsid Protein Purification from *E. coli*

With the *lac* promoter in the BL21(DE3) cells, MSV CP can be expressed with IPTG induction. The overnight IPTG-induced *E. coli* were harvested by centrifugation at 6,000g at 4°C for 30 minutes. Cells were re-suspended in lysis buffer (20 mM HEPES [pH 7.5], 100 mM NaCl). The cells were lysed by sonication for 5 minutes (5 s pulses with 5 s gaps) on ice using a Vibra-CellTM (Sonics). The whole-cell lysates were cleaned by centrifugation at 40,000g for 40 minutes to remove cell debris. The resulting supernatant was transferred to 5 mL of pre-equilibrated Co²⁺ charged Talon metal affinity beads (Clontech) and incubated for 1 hour at 4°C with gentle mixing. Beads with bound protein were collected and washed with 10 ml of wash buffer (5 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole) three times. The protein was then eluted with 10 ml elution buffer (5 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole). The eluted sample was analyzed by SDS-PAGE and further purified with size exclusion chromatography (Superdex S200, GE Healthcare) at a flow speed of 0.5 ml/min in gel filtration buffer (100 mM NaCl, 20 mM HEPES pH 7.5). Peak fractions containing target protein were pooled and concentrated to 1 mL.

MSV-CP was further purified using ultracentrifugation. Sucrose gradients (10%-40%) were made with The Gradient Master (BIOCOMP) and sample was loaded on top of the gradient. Ultracentrifugation was carried out for 3 hours at 4°C at 210,000 g (Beckman Coulter, OptimaTM XPN) before the light scattering layer containing virus-like particles was collected. Samples were further analysed with transmission electron microscopy.

2.6. Protein Expression in insect cells (*Sf9* cells)

2.6.1. *Sf9* Cell Culture

Sf9 cells were cultured in Sf900II medium (Invitrogen) and maintained according to the standard OPPF protocol (<https://www.oppf.rc-harwell.ac.uk/OPPF/protocols>) Cells were split to 4×10^5 cells/ml every 3-4 days, and incubated with gentle shaking at 27°C.

2.6.2. Generation of recombinant baculoviruses

An aliquot (2ml) of *Sf9* cells at 6×10^5 cells/ml were seeded to each well of a 6-well plate and allowed to attach for 10 minutes. To transfect the *Sf9* cells, 3 µl FugeneHD transfection reagent (Roche applied science) and DNA mixture (1.5 µl bacmid and 6 µl plasmid DNA) were first mixed with 100 µl of Sf900II medium in separate Eppendorf tubes before being combined and incubated for 20 minutes at room temperature (Zhao et al. 2003). The mixtures were then further diluted with 800 µl Sf900II medium. The medium from each well was removed and replaced with 1 ml transfection mixture from above. The cells were incubated at 27 °C in a static tissue culture incubator for 24 hours before another 1 ml of Sf900II medium was added to each well. The viruses were left to propagate for 6 days before harvesting. The P1 virus stock was harvested and stored at 4°C (Aricescu & Assenberg 2006).

2.6.3. Preparation of Stock Baculoviruses

Tissue culture flasks (175 cm²) were seeded to 80% confluency with *Sf9* cells ($\sim 38\text{--}40 \times 10^6$ /flask). Cells were allowed to attach to the flask for 15 minutes before 100 µl of P1 virus was added. The infected cells were incubated for 6 days before the medium was harvested as P2 virus stock.

2.6.4. BTV15 VP5 Protein Expression

All recombinant proteins were tested in small-scale expression experiments (6-well plates) before large-scale expression for crystallization. Tissue culture flasks (175 cm²) were seeded to 80% confluency with *Sf9* cells (~38-40x 10⁶/flask). Cells were allowed to attach for 15 minutes before 1 ml of P2 virus was added. The infected cells were incubated for 3 days before harvesting by knocking the flasks and detached using a cell scraper.

2.6.5. Selenomethionine-incorporated protein expression

Sf9 cells were seeded to tissue culture flasks (175 cm²) with 80% confluency (~38-40 x10⁶/flask) and allowed to attach for 15 minutes before 1 ml of P2 virus were added. 20 hours post-infection, the cell media were replaced with cysteine- and methionine-free SF900II SFM media supplemented with dialyzed FBS to 10% (v/v) and 150 mg/L L-cysteine. Cells were allowed to grow for 4 hours to deplete the cellular methionine level before a stock of Selenomethionine (SeMet) was added to the culture to give a final concentration of 250 mg/L. The expression was continued for another 2 days for optimal levels of expression.

2.7. Protein purification from *Sf9* cells

2.7.1. Cell harvest and homogenizing

Cells were pelleted by centrifuging at 3,000 x g for 30 minutes at 4°C and the supernatant discarded. The pellet was then re-suspended in 5 mM HEPES buffer (pH 7.5). Cells were lysed as follows: incubation on ice for 10 minutes, 20 strokes of pre-cooled Dounce homogenizer, and 10 minutes on ice. The lysates were adjusted to 500

mM NaCl by adding a 5M stock followed by centrifugation at 40,000 x g for 30 minutes at 4°C to remove cell debris.

2.7.2. Primary purification by binding to Cobalt beads

The supernatant was incubated with 3 mls of Co²⁺ charged Talon metal affinity beads (Clontech) per 20 flasks of cells for 1 hour at 4°C with gentle mixing. Beads with protein bound were collected, washed with 10 mls of wash buffer (5 mM HEPES pH 7.5, 500 mM NaCl, 10 mM Imidazole) three times. The protein was then eluted with five aliquots of 3 ml elution buffer (5 mM HEPES pH 7.5, 500 mM NaCl, 250 mM Imidazole). The protein solution was then buffer-exchanged into 5 mM HEPES pH 7.5, 500 mM NaCl with a PD10 desalting column (GE Healthcare).

2.8. Purification of maize streak virus from plants

Maize seedlings were infected by agro-inoculation, which is an *Agrobacterium*-mediated DNA transfer (Grimsley et al. 1989; Boulton et al. 1989). In this method, sweetcorn kernels (Golden Bantam) were germinated and grown for three days at 30°C before seedlings were cleaned and injected with a solution of *agrobacterium* cells at a site near the plant meristem. An obvious signs of successful infection is streaky leaves which can be observed from the third week. The plants were harvested 4 weeks post-infection, cut into 2-3cm pieces and frozen at -20°C until further use. Only leaves were collected rather than including stems since stems result in more contaminants. Leaves (10g) were put into a beaker with 50 ml buffer (100 mM sodium acetate pH 4.5, 100 mM NaCl), and the mixture was blended for 5 min to make sure the leaves were thoroughly broken. The blended mixture was filtered through double layer cloth and centrifuged at 7,000g for 40 minutes at 8°C. The supernatant was further centrifuged

by ultracentrifugation carried out at 8°C, 130,000g for 2.5 hours to pellet the virus particles. The pellet was re-suspended gently with a small volume of buffer (100 mM sodium acetate pH 4.5, 100mM NaCl) followed by a short centrifugation at 5,000g for 1 minute to remove the insoluble material. The supernatant containing the maize streak viruses was checked with negative stain electron microscopy for quality.

2.9. Crystallization

2.9.1. Crystallization Screens

Before setting up crystallization trials, a pre-crystallization test (PCT, Hampton Research) was performed initially for each protein sample to determine an appropriate protein concentration. Crystallization trials were performed in a sitting drop vapour diffusion format using 100nL protein solution added onto 100nL reservoir solution set up in 96-well Greiner plates with Cartesian Technologies robots (Walter et al. 2005). Dispensed plates were either kept in a Formulatrix Rock Imager 1000 storage vault at 20°C and imaged via a Veeco visualisation system or kept in a storage vault at 6.5°C. The standard range of crystallization screens used for each new protein construct at both 20°C and 6.5°C are listed in Table 2.3.

Table 2.3 Standard crystallization blocks used in this study.

Masterblock Screen	Wells	Commercial Screen	Manufacturer
Block 1	A1-D12	Crystal Screen	Hampton Research
	E1-H12	Crystal Screen 2	Hampton Research
Block 2	A1-D12	Wizard 1	Emerald Biosystems
	E1-H12	Wizard 2	Emerald Biosystems
Block 3	A1-D12	PEG/Ion screen	Hampton Research
	E1-F12	PEG 600 grid screen	Hampton Research
	G1-H12	Ammonium sulphate screen	Hampton Research
Block 4	A1-D12	Natrix screen	Hampton Research
	E1-H12	Crystal screen cryo	Hampton Research
Block 5	A1-B12	PEG/LiCl grid screen	Hampton Research
	C1-D12	NaCl grid screen	Hampton Research
	E1-F12	MPD grid screen	Hampton Research
	G1-H12	Quik screen	Hampton Research
Index	A1-H12	Index Screen	Hampton Research
Salt-Rx	A1-H12	Salt-Rx screen	Hampton Research
PACT premier	A1-H12	PACT premier	Molecular Dimensions
PEG-Rx	A1-H12	PEGRx screen	Hampton Research
PentaErythritol	A1-H12	JBScreen Pentaerythritol	Jena Bioscience
Morpheus	A1-H12	Morpheus screen	Molecular Dimensions
ProPlex	A1-H12	Proplex screen	Molecular Dimensions
PGA	A1-H12	The PGA screen	Molecular Dimensions

2.9.2. Crystal Optimizations

Crystals formed from the initial screening experiments were routinely optimized with two types of optimization procedures. The first of these is adjusting the percentage of the components and pH in the reservoir with either sequential dilution or proportional adjustment. Different drop ratios of protein solution to reservoir solution (normally 1:1, 1:2, 2:1) can be applied at the same time. The second optimization procedure commonly used was the Hampton additive screen. The 96-well crystallization plates were set up using protein solution and well solution, then 100 nL of the additive screen reagent was added directly to the corresponding drops.

2.10. X-ray crystallography data collection and processing

2.11. Cryo-protection and crystal mounting

In order to decrease the extent of crystal deterioration by radiation damage, protein crystals were protected in cryogenic conditions before X-ray data collection. Crystals were fished from the 96-well Greiner plates using cryo-loops matching the size of the crystal. Initially as potential cryo-protectants, 25-30% glycerol or 25-30% ethylene glycol were tested for all protein crystals. The cryo solution was added directly into the drop if single crystals were present in a drop. Otherwise, crystals were fished from the mother liquor solution and then submerged into cryo-protectants before promptly being moved into liquid nitrogen and stored until data collection. 30% ethylene glycol was used as the cryo-protectant for the native BTV15 VP5 crystals and 25% glycerol was used for the SeMet form of BTV15 VP5 protein.

2.11.1. Data collection

X-ray diffraction data were collected on beamlines at Diamond Light Source (DLS), Oxfordshire, UK. Prior to each data collection, an appropriate data collection strategy was determined for each crystal with the `fast_dp` and `mosfilm` (Battye et al. 2011) program built into the data collection interface at DLS. Complete datasets were collected after the optimum orientation was found. Data were collected with Pilatus detectors in cryogenic conditions at 100K.

The datasets for BTV15 VP5 native crystals were collected on beamline I24 with a Pilatus3 6M detector. Two datasets of 180 degrees were collected with a 20 μm x 20 μm beam and with an oscillation of 0.05°. A total of 3600 images were taken for each of the two datasets with 0.02 seconds of exposure time per image.

The data for BTV15 VP5 SeMet crystals were collected on beamline I03 with a Pilatus3 6M detector. Prior to the data collection, a fluorescence scan was carried out to confirm the presence of selenium and the location of the absorption edge. Five datasets of 180 degrees were collected with a 50 μm x 20 μm beam and with an oscillation of 0.1°. Two datasets, each of 1800 images, were collected with 0.02 seconds exposure time per image.

2.11.2. Data processing

Data processing includes continuous procedures of indexing into the right space group and cell dimensions, integration of datasets, and scaling of the reflection intensities from the entire dataset. For this purpose, there are data processing suites like HKL2000 (Zbyszek & Wladek 1997), Mosfilm (Leslie & Powell 2007), and several versions of xia2 pipeline programs to use (Evans 2006; Sauter et al. 2004). Data quality were assessed with several criteria such as completeness (normally > 90%), which is the

percentage of expected unique reflections that are measured in the dataset; redundancy (aim for higher), the ratio of total measured reflections to unique measured reflections in the dataset; $I/\sigma I$ (normally > 1.5 at high resolution), the ratio of signal intensity to the error in measurement of the signal intensity; and R_{merge} (aim for lower than 30%), the measurement of internal consistency of the dataset.

For BTV15 VP5 native data, two sets of data were merged together and processed using the xia2 3dii pipeline which uses XDS and XSCALE (Kabsch 2010b; Kabsch 2010a; Kabsch 1988) to index with peaks found from all images. For BTV15 VP5 SeMet data, the five datasets collected were merged and processed with HKL2000, to produce a data file.

2.12. Crystal Structure determination

2.12.1. Phasing by molecular replacement

It is the intensities of the diffracted waves scattered from series of planes that we are collecting from X-ray diffraction of a crystal. From these intensities we can derive the amplitudes of the scattered waves, but we lose the phase information which is how we offset the waves when we add them together to reconstruct the image of molecules. In protein crystallography, phases are determined mainly in two ways: the first is by finding the positions of heavy atoms which are either intrinsic to the protein or which have been added to it. The second way is by using the atomic coordinates of a protein of similar structure, a method called molecular replacement (Mccoy et al. 2007). The matthew_coef program (Kantardjieff 2003; Weichenberger & Rupp 2014; Matthews 1968) in the CCP4 suite (Winn et al. 2011) can provide an initial estimation of the number of protein molecules in an asymmetric unit. The Molrep program (Vagin 2010)

in the CCP4 suite was used to find an appropriate number of copies of models in two steps: 1. Rotation function, to search for the correct orientation of the model and 2. Cross translation function and packing function, to search for the correct position of the oriented model. For both of the high pH SeMet BTV15 VP5 and low pH native BTV15 VP5 proteins, models derived from the structure of the virus (PDB code: 3J9E) were used as search models for molecular replacement phasing. For the high pH SeMet BTV15 VP5 dataset, a N-terminal 43 amino acid truncated form of PDB_3J9E was used as a search model while for the low PH native BTV15 VP5, the anchor domain of PDB_3J9E (amino acid number of 355-526) was used for the search model. For the high pH structure, the phasing was also confirmed using selenium methionine positions, processed with the program Anode (Thorn & Sheldrick 2011).

2.12.2. Refinement

From the initial molecular replacement model obtained using the BTV1-VP5 EM structure, the amino acids sequence was corrected in COOT (Emsley & Lohkamp 2010) with the single mutation tool, chains directions were traced and corrected according to the electron density map. Refmac5 (Skubak et al. 2004) was used to refine the model using restrained refinement before models were further adjusted with real-space and geometry restrained refinement in COOT with appropriate geometry weights. For the low pH native BTV15 VP5 structure, a long helix of amino acids between 68 and 140 could be built in manually after several rounds of refinement. For the high pH SeMet BTV15 VP5 structure, a round of global refinement in Refmac5 with applied NCS restraints was used before a round of local refinement treating every molecule in the asymmetric unit separately was used and all models were manually corrected.

2.12.3. Structure validation

Structure validation is the comparison of various aspects of the model with pre-conceived notions of “good quality”. For structure validation, three in-built functions in COOT were used to assess the stereochemistry properties of the models. Geometry analysis compares the geometry (bonds, angles, planes) for each residue in the model against reference values. The second validation tool is Ramachandran Plot, which is a way to visualize energetically allowed regions for backbone dihedral angles ψ against ϕ of amino acid residues (Ramachandran et al. 1963), ideally there should be less than 1% outliers. Thirdly the rotamers were checked using the function in COOT to correct the orientation of the side chains since the side-chains have certain preferred combinations of torsions around their rotatable bonds.

2.13. Transmission electron microscopy data collection

2.13.1. Basic principles of negative stain electron microscopy (EM)

To visualize the structure of proteins or viruses, negative stain electron microscopy (EM) is a fast and convenient way in which the target object is originally freely diffused in a solution environment and during the subsequent sample preparation embedded into a layer of heavy metal stain, normally either uranyl acetate or uranyl formate in this lab. The use of a heavy metal stain can provide high image contrast by the uranyl atoms surrounding the sample. Heavy atoms absorb electrons significantly more than the light atoms making up the sample while also protecting the specimen from electron bombardment. Negative stain EM was used as a complimentary technique to protein crystallography since it may reveal the natural structural arrangements otherwise obstructed by constraints in crystal packing although it should

be noted that during the drying in sample preparation may introduce artefacts into the sample. In recent years, the use of cryo-EM has become more widespread and powerful, so that negative stain- EM is less used for structural determination but more in initial sample checking and determination.

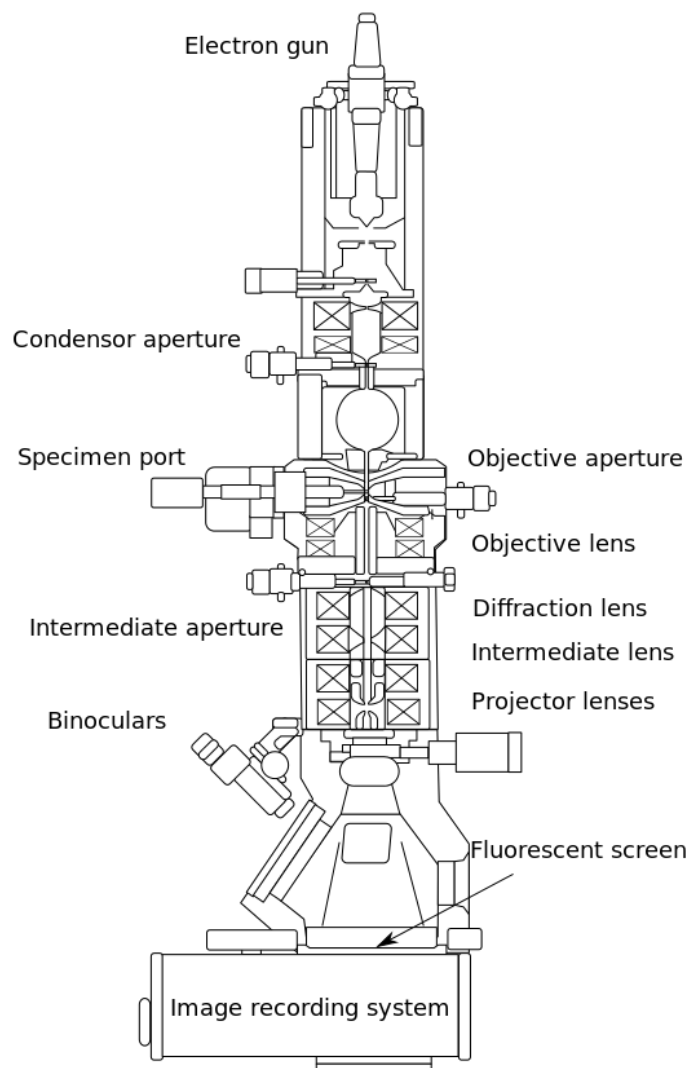


Figure 2.2 Schematic graph of a transmission electron microscope.

The basic components of a transmission electron microscope (TEM) include an electron source, a series of electromagnetic lenses, apertures, and image detection

systems (Figure 2.2). A small portion of electrons are absorbed by the heavy metal atoms surrounding the specimen when electrons strike a negative stained sample resulting in reduced intensity at stain-ingrained regions in the image thus generating an amplitude contrast. In negative stain EM the EM amplitude contrast contributes the majority of image contrast while only a very small fraction of electrons collected by the detector are scattered elastically and interfere with the non-scattered electrons resulting in phase contrast. In practice, image contrast can be increased by adjusting the projector lenses to obtain larger defocus values, but at the same time it may sacrifice resolution thus need to be carefully adjusted in a case-by-case manner.

2.13.2. Sample preparation for negative stain EM

A four-drop method was used to fix the sample onto a carbon-coated grid and stain it. A drop of 4 μ L freshly prepared sample was absorbed to the newly glow-discharged carbon-coated copper grid, left for 30 seconds and then dried with filter paper. The grid was washed with two drops of 50 μ L deionized water continuously, dried with filter paper each time, followed by staining the sample with two drops of 20 μ L uranyl formate or uranyl acetate heavy-metal stain, excess stain was carefully removed after each time before air-drying.

2.13.3. Sample checking on Tecnai T12 Electron Microscopy

Liquid nitrogen was added to the cold trap and it took at least 10–15 min to reach its optimum temperature before turning on the “high tension” and “filament”. The specimen grid was inserted into the column and the pump out was allowed to finish, (i.e. red light turns off) before the holder was rotated counter-clockwise about 120°, until the column vacuum pulled the holder in. The specimen was placed at the eucentric height and focus was then adjusted by activating the alpha wobbler to adjust

Z height at a relatively low (1000–5000x) magnification. Under a higher magnification, the focus and beam intensity were adjusted again slightly to gain an optimized view of the specimen before the phosphor screen was lifted so the beam can reach the camera. The specimen was viewed and further focused in real time in “search” mode while “preview” mode gave a preview of higher resolution at the final acquired image quality but only uses half of the CCD area. When the preview image reached the required quality, “acquire” was used to capture a single image from the whole CCD detector. Data were acquired on a 4k x 4k FEI Eagle CCD-camera at a magnification of 67,000x, corresponding to a pixel size of 1.68 Å/pixel.

2.13.4. Graphene oxide grid preparation

The graphene oxide (GO) stock solution was diluted by ten times to 0.2 mg/mL with Milli-Q water (MQ H₂O) and gently centrifuged for a minute at 300 rpm to remove aggregates. Holey carbon-coated C-flat grids (C-flat, CF-2/1-2C, Protochips) were glow discharged with carbon side up while a clean worktop covered with parafilm was prepared. For each grid 3 x 20 µL drops of ddH₂O were prepared. To coat grids with GO, 3 µL of the diluted GO solution was placed onto a grid and incubated for 1 minute before it was removed with filter paper. The grid carbon side was washed with MQ H₂O twice, and dried after each wash with filter paper. The third wash was carried out in the same way, except that only the back side of the grid was washed. Grids were used after being air dried for 5 minutes.

2.13.5. Sample preparation for cryo-EM with Vitrobot

The Vitrobot sample freezing cup is cylindrical and consists of a central cup for liquid ethane into which the grid is plunged and a larger peripheral ring for liquid nitrogen. Between these chambers is a gap spanned by carefully designed thermal connectors

that conduct just enough heat that the ethane will first liquefy and then very slowly freeze. To prepare liquid ethane cooled with liquid nitrogen, the central cup together with the thermal connector were assembled accordingly. The outer nitrogen ring was then filled with liquid nitrogen to cool down the central cup. After the central cup was cooled down (the liquid nitrogen in the outer ring stopped bubbling), ethane gas with a modest flow rate was directed towards the bottom or side of the cold cup wall. The gaseous ethane was slowly liquefied in the cup and the thermal connector was removed after the ethane cup was filled with ethane. Grid boxes were placed in their holders in the outer ring at this time and cooled along with the Vitrobot cup. The liquid nitrogen in the outer ring was maintained at least 75% full at all times. All the previous procedures were carried out in a fume hood.

After the liquefied ethane was prepared, the Vitrobot was prepared by turning on the tank of compressed air, ensuring that the proper air pressure (at least 6 bar) is applied. The humidifier was then filled with distilled water and the filter paper was mounted on the blotting pads with the clip rings secured on them. The Vitrobot was set with the desired parameters in the Console and Options screens: temperature, humidity (100%), type of application set as automatic. A grid was picked up by its edge with the Vitrobot tweezers and mounted onto the tip of the plunger. After the blotting chamber reached the desired temperature and humidity, the plunger was activated and will automatically ask for sample application, followed by blotting and plunging in the ethane. After the plunge, the Vitrobot tweezers were slide off the tip of the plunger while keeping the grid submerged within the liquid ethane, and transferred into the grid box and stored in liquid nitrogen until further use.

2.14. Data collection, processing and refinement for maize streak viruses for cryo-EM

2.14.1. Cryo-EM data collection

Data collection was performed with the Tecnai F30 Polara electron microscope (FEI) based in Division of Structural Biology, University of Oxford. Transfer stations were used to keep all cryo-EM grids under liquid nitrogen when loading grids and changing between grids. A multi-specimen rod was used to load grids into the microscope column. The specimen was first brought to eucentric height by adjusting the position of the sample on the z-axis after centring the beam and checking for alignments, before imaging took place. The grid was scanned in Search mode and areas of interest were identified at low-magnification, which were then marked for data collection. On the Tecnai F30 Polara electron microscope (FEI), the cryo-EM grids were imaged using a Direct Electron Detector (K2 Summit, Gatan Inc.) at 300 kV and equipped with an energy filter (GIF Quantum, Gatan) operated in zero-loss mode (0-20 eV energy slit). For each movie taken, 35 frames were recorded for 0.2s. The defocus was changed by 0.1 μm steps between movies to cover a range of -3.5 to -1.5 μm . Data collection was performed using SerialEM with the detector operating in counting-mode at a nominal magnification of 37,027x with a resultant pixel size of 1.35 $\text{\AA}/\text{pixel}$, and an electron dosage of 25 electrons per $\text{\AA}^2$.

2.14.2. Basic principles of cryo-EM data processing in Relion

The collected movie frames were first aligned and motion corrected with MotionCorr (Li et al. 2013) to produce high quality micrographs. Micrographs with significant astigmatism and large drift caused by the electrons interacting with the specimen were removed from subsequent analyses. Normally, to exclude the effects of beam-induced

drift from the initial exposure and accumulated radiation damage, the first two and the last frames from each movie were excluded from the alignment.

The next step in data processing is the determination of the Contrast Transfer Function (CTF). The CTF represents how much contrast from each Fourier component of each electron is transferred to the final image. This information can be depicted using a power spectrum of the image: a set of concentric Thon rings showing the alternating pattern of the presence (the maxima in a CTF function) and absence of signal (the zeros in the CTF function) for the entire exposure that produced the image. The centre of the spectrum represents low-resolution information and circles increasingly further out circles depict the high-resolution details. Since the CTF is different for different defocus values, a different range of contributing spatial frequencies was recorded at each defocus by imaging a specimen at a range of defocus values, and both low and high resolution features from the sample are imaged for the final structure by collecting data at several defocus values. To perform the CTF correction, the Relion software suite (Scheres 2012) was used. The CTFFIND algorithm (Mindell & Grigorieff 2003) in the Relion suite was run to match the theoretical CTF to the experimental values for CTF correction and the micrographs were manually inspected. Micrographs with astigmatism, large drifts or strange diffraction patterns due to large ice crystal contamination were removed, leaving those micrographs with a well-fitted CTF to be carried on to the subsequent processing.

For the micrographs selected, individual particles were extracted from the micrographs. Particle picking was performed using e2boxer.py from the EMAN2 software suite (Tang et al. 2007) with which the particles were picked either manually or automatically. The coordinates of the picked particles were saved as *.box files, from which the particles can be later extracted by the Relion software suite with a given box

size. When the particles were automatically picked, a manual screening was performed to ensure no apparently deformed particles or contaminants were included.

The particles contained in one grid may be frozen in several different orientations, and these different views of the particle in a number of micrographs were combined to produce the final structure. These different particles were aligned with Relion with the Maximum Likelihood (ML) approach, by allowing x- and y-shifts as well as in-plane rotations, resulting in homogenous 2D classes. The final classes were checked manually by visual inspection and based on the angular-rotation accuracy obtained from the model.star file in Relion. After the 2D projection images of the specimen were classified, they were sorted into homogenous populations and reconstructed into a 3D structure. This process is based on the Fourier slice theorem: the Fourier transform of each 2D projection is a centre-slice of the Fourier transform of the 3D structure of the specimen, and the inverse Fourier transform generates the 2D-projection of the respective 3D density perpendicular to the slice in Fourier space. In this sense, each 2D projection image contributes to the 3D representation of the specimen in Fourier space, and the more 2D projections of the sample are acquired (more particles in different views), the finer the 3D Fourier space sampled and the better the final 3D reconstruction.

An initial model at low resolution was needed to start this iterative process without generating reference-bias. For the MSV-VLP reconstruction, a 7 Å crystal structure gained earlier was filtered to 50 Å resolution and used as an initial model, while for the geminated MSV virus reconstruction, a model was made by manually joining two incomplete icosahedrons together. Projections in 2D were generated from the 3D model to create references before the particles were then cross-correlated to each of the template 2D projections in Relion in an iterative way. During each iteration, an in-

plane rotation and an x-, y-shift was used to match each particle to the model and for each iteration a new and progressively better model was generated until it converged on a consensus 3D model of the structure. In refinement, the entire dataset was split into two halves and the 3D structural refinement performed for each half independently. For each half, Fourier transformation was carried out and correlated against the reconstruction from the other half. The correlation is plotted against resolution to see how good the match is between the two halves. This is termed the “Gold Standard” refinement procedure.

Post-processing using Relion (Chen et al. 2013) was also used to further refine the map. It was used after a 3D refinement calculation for automated masking, MTF-correction and B-factor sharpening. Because the Fourier Shell Correlation (FSC) calculations inside the 3D auto-refine procedure depend on unmasked maps (to avoid potential overfitting), the auto-refine procedure may somewhat under-estimate the true resolution. By filtering the map, top-quality reconstructions may be obtained (Chen et al. 2013).

2.14.3. Localized reconstruction on two halves of geminated MSV viruses

Electron cryo-microscopy can yield near-atomic resolution structures of highly ordered macromolecular complexes. However, the attainable resolution may be limited when some subunits bind in a flexible manner, or are present in sub-stoichiometric amounts. At this time, localized three-dimensional reconstruction of sub-particles extracted from the complex and treated as new single particles can be used to optimize the data processing. These sub-particles were created by describing a vector showing the position of one sub-particle relative to the particle origin. The localized script was fed with both the vector and the symmetry of the sub-particles and

then the sub-particles can be extracted. In the case of geminated MSV particles, the sub-particles were defined as both halves of the particles, each contain one incomplete icosahedron. After the sub-particles were extracted, further processing follows that of single particles (Ilca et al. 2015).

2.15. Multi-angle light scattering analysis

Multi-Angle Light Scattering (MALS) experiments were carried out using an analytical Superdex S200 10/30 column (GE Healthcare) with online static light scattering (DAWN HELEOS II, Wyatt Technology, Santa Barbara, CA), differential refractive index (Optilab rEX, Wyatt Technology) and Agilent 1200 UV (Agilent Technologies) detectors. Proteins used for MALS had previously been purified by size-exclusion chromatography and concentrated to approximately 2 mg/ml for BTV15 VP5 protein. Data were analyzed using the ASTRA software package (Wyatt Technology).

Chapter 3.

Results of BTV VP5 structures

3.1. Expression constructs screen for bluetongue virus VP5

One of the major challenges in protein crystallization is to find constructs showing highly expressed, soluble and crystallisable proteins. Often it is necessary to screen a lot of constructs from different serotypes of viruses. There are 26 serotypes of BTV. In order to pick up representative serotypes from bluetongue virus, sequence alignment tools were used to group the VP5 proteins from all serotypes of BTV. Protein sequences were aligned using the online program ClustalW2 (EMBL-EBI) and displayed with the online program ESPript (esprict.ibcp.fr/). The alignment result shows that the amino acid sequences of VP5 proteins among serotypes are in general rather highly conserved with some segments showing more significant variation (Figure 3.1). The alignment was used to construct a phylogenetic tree and the numbers shown on the branches marked the distance in sequence similarity (Figure 3.2A). VP5 proteins from 26 serotypes can be grouped into 8 groups labelled A-H according to the phylogeny (Figure 3.2B). This phylogenetic grouping provides good guide for picking representative serotypes to design protein expression constructs. About one serotype was chosen from each group for subsequent cloning and expression screen.

93

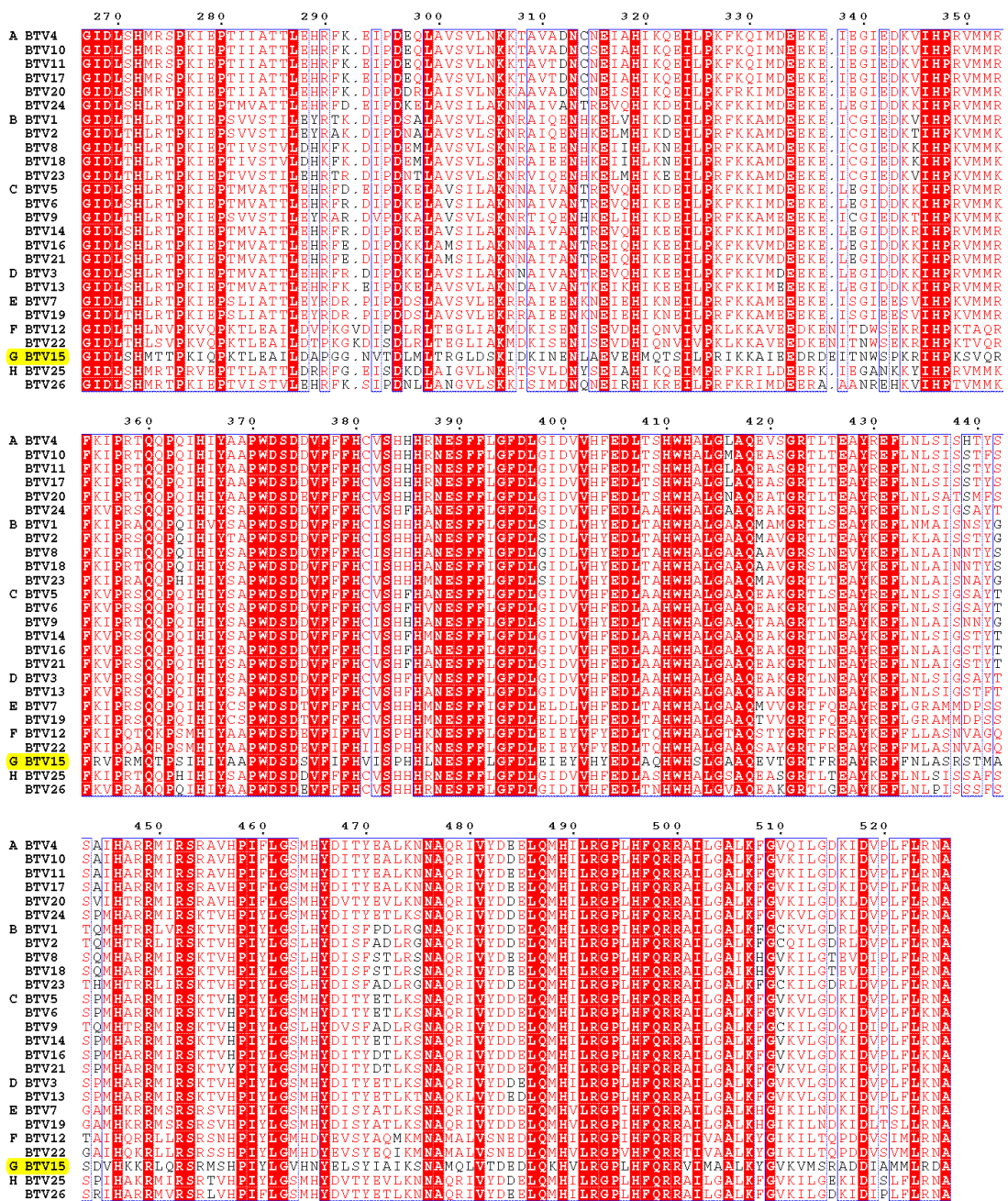


Figure 3.1 Sequence alignments of 26 serotypes of BTV. VP5 protein sequences were aligned based on Risler Matrices (Liebert et al. 2001). Red box with white characters indicates strict identity among sequences; red characters show similarity in a group; blue frame shows similarity across groups. The sequences were grouped (A to H) according to the phylogenetic analysis in Figure 3.2, with BTV-15 highlighted at the beginning of the lines.

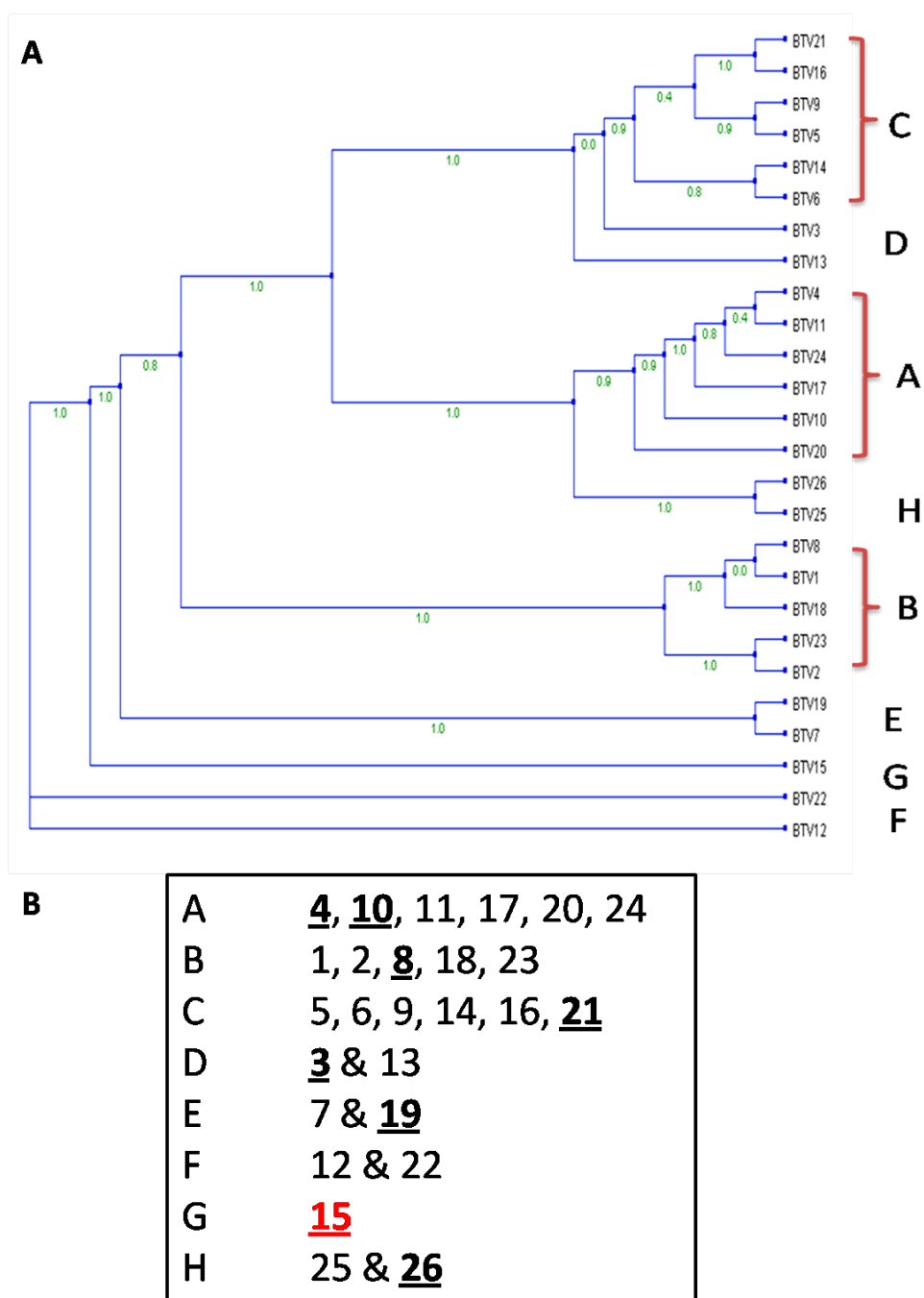


Figure 3.2 Phylogenetic grouping bluetongue virus VP5. (A) Phylogenetic tree generated from protein sequence alignment. (B) VP5 proteins from 26 serotypes can be grouped into 8 groups named A-H according to the sequence alignment. The serotypes in bold letters and underlined are the ones which constructs were tested for expression and solubility. The serotype in red colour is the one used for crystallization experiment and further structural study.

3.2. Protein expression and purification

3.2.1. Baculovirus Expression system

In the protein crystallization field, protein expression in bacteria is still the most popular method. However, more challenging targets, such as viral and human proteins are often poorly expressed in bacteria. Because of this, eukaryotic expression systems using insect or mammalian cells are becoming increasingly popular in dealing with “difficult” targets. Among the eukaryotic expression systems, the preferred system for producing non-secreted proteins for structural studies is the baculovirus *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) expression system to produce recombinant proteins in insect cells. The polyhedrin gene promoter is the strongest promoter in eukaryotic cells, and is therefore generally used to ensure high levels of recombinant protein expression. As the AcMNPV genome is quite large (about 130 kB), direct cloning of recombinant genes is difficult. Recombinant virus is usually generated by co-transfecting insect cells with a transfer vector and linearized baculovirus genomic DNA. The transfer vector pOPINE or pOPINF (Berrow et al. 2007) was used to clone in bluetongue VP5 constructs and generate recombinant baculovirus. The essential gene (ORF1629) knocked out baculovirus genomic DNA (Zhao et al. 2003) was used for co-transfection.

3.2.2. Construct design and cloning

The baculovirus transfer vector pOPIN series created by the Oxford Protein Production Facility (OPPF) was used for cloning of BTV VP5 gene with In-Fusion based ligation-independent cloning (LIC). The pOPINE and pOPINF vectors used here to facilitate his tagging at either N- or C-terminal of the VP5. Both vectors are similar, and a detailed vector map of the pOPINE is shown in Figure 3.3. pOPINF is very much the

same as pOPINE, the only difference is the tag usage. pOPINE has a KHHHHHHH C-terminal tag, while the pOPINF has a 3C protease cleavable tag with sequence MAHHHHHHSSG-LEVLFQGP at the N-terminus.

The full-length VP5 and two N-terminal truncations of 21 and 41/43 amino acids were used for all the chosen serotypes for the expression screen. The N-terminal truncations were designed to remove either 1 or 2 amphipathic α -helices which have been shown to have a deleterious effect on protein expression. The set of constructs is detailed in Table 3.1 with the OPPF construct number, construct name, vector used, gene size and protein size for each construct. The DNA for the constructs was amplified by PCR from the synthesized full-length VP5 gene (GeneArt). Genes for different constructs were then cloned into the appropriate pOPIN vectors using the In-Fusion system.

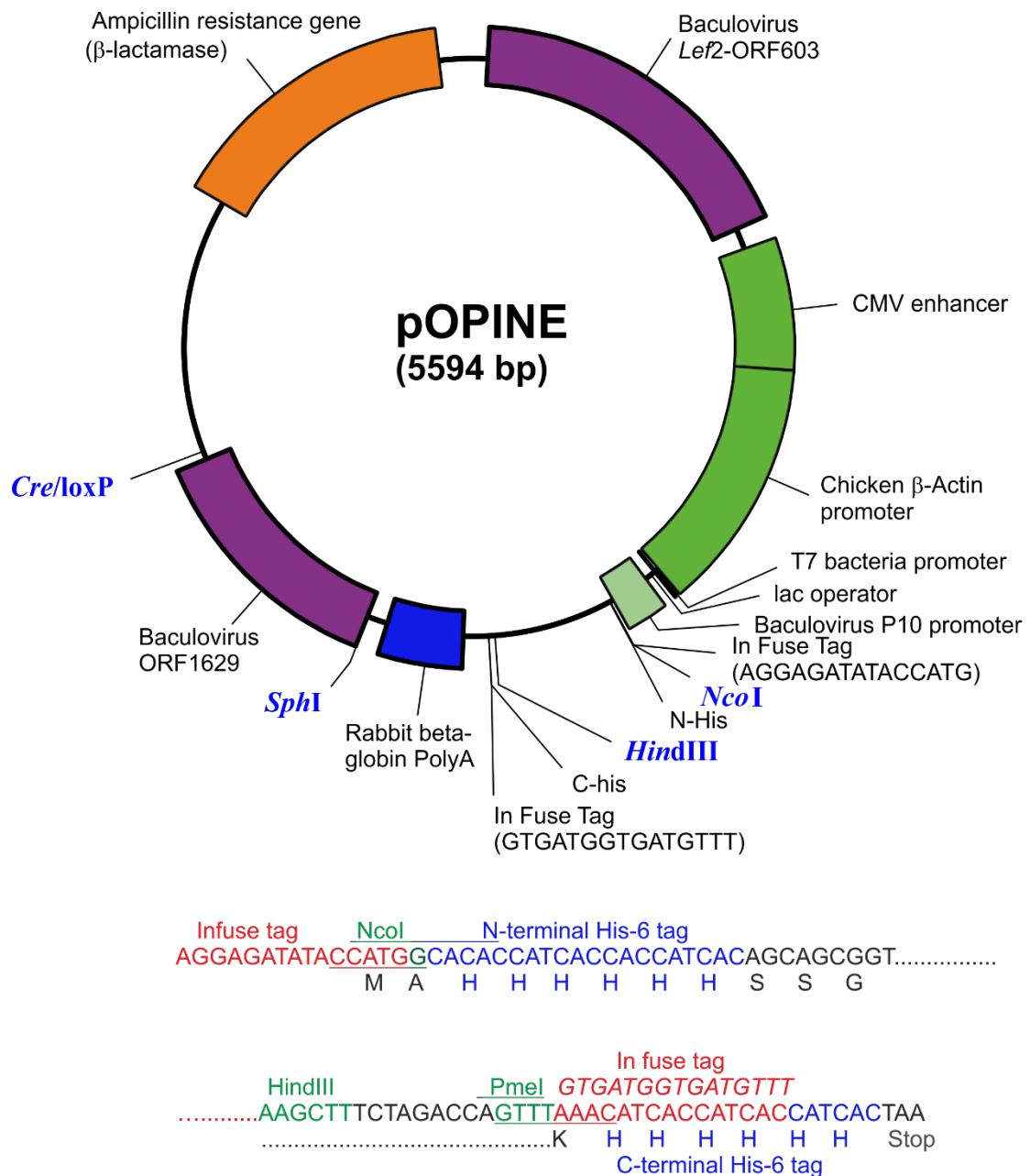


Figure 3.3 The pOPINE baculovirus transfer vector map. The baculovirus sequences used for homologous recombination are coloured purple. Promoters for expression in mammalian cells, baculovirus insect cells, and bacteria are coloured green with labelling. In Fusion tag sequences are showing at the bottom.

Table 3.1 Detail cloning information about BTV VP5 expression constructs.

OPPF No.	CONSTRUCT	VECTOR	PCR SIZE (bp)	PROTEIN SIZE (kDa)
8026	BTV10 VP5 FL N his	pOPINF	1616	61.2
8027	BTV10 VP5 FL C his	pOPINE	1608	60.2
8030	BTV10 VP5 43 N his	pOPINF	1455	56.8
8031	BTV10 VP5 43 C his	pOPINE	1452	55.7
8183	BTV10 VP5 21 N his	pOPINF	1556	59.2
8184	BTV10 VP5 21 C his	pOPINE	1548	58.1
8028	BTV15 VP5 FL N his	pOPINF	1619	60.8
8029	BTV15 VP5 FL C his	pOPINE	1611	59.9
8032	BTV15 VP5 43 N his	pOPINF	1496	56.6
8033	BTV15 VP5 43 C his	pOPINE	1485	55.5
8185	BTV15 VP5 21 N his	pOPINF	1559	58.9
8186	BTV15 VP5 21 C his	pOPINE	1551	57.9
14596	BTV3 VP5 FL C his	pOPINE	1608	60.3
14597	BTV3 VP5 FL N his	pOPINF	1613	61.2
14598	BTV3 VP5 21 C his	pOPINE	1545	58.0
14599	BTV3 VP5 21 N his	pOPINF	1553	59.1
14600	BTV3 VP5 41 C his	pOPINE	1485	55.9
14601	BTV3 VP5 41 N his	pOPINF	1493	56.9
14602	BTV4 VP5 FL C his	pOPINE	1605	60.2
14603	BTV4 VP5 FL N his	pOPINF	1613	61.3
14604	BTV4 VP5 21 C his	pOPINE	1545	58.2
14605	BTV4 VP5 21 N his	pOPINF	1553	59.2
14606	BTV4 VP5 41 C his	pOPINE	1485	56.0
14607	BTV4 VP5 41 N his	pOPINF	1493	57.1
14608	BTV8 VP5 FL C his	pOPINE	1605	60.3

14609	BTv8 VP5 FL C his	pOPINF	1613	61.4
14610	BTv8 VP5 FL C his	pOPINE	1545	58.2
14611	BTv8 VP5 FL C his	pOPINF	1553	59.3
14612	BTv8 VP5 FL C his	pOPINE	1485	56.0
14613	BTv8 VP5 FL C his	pOPINF	1493	57.1
14614	BTv19 VP5 FL C his	pOPINE	1605	60.4
14615	BTv19 VP5 FL C his	pOPINF	1613	61.4
14616	BTv19 VP5 FL C his	pOPINE	1545	58.3
14617	BTv19 VP5 FL C his	pOPINF	1553	59.3
14618	BTv19 VP5 FL C his	pOPINE	1485	56.1
14619	BTv19 VP5 FL C his	pOPINF	1493	57.1
14620	BTv21 VP5 FL C his	pOPINE	1605	60.2
14621	BTv21 VP5 FL C his	pOPINF	1613	61.2
14622	BTv21 VP5 FL C his	pOPINE	1545	58.1
14623	BTv21 VP5 FL C his	pOPINF	1553	59.1
14624	BTv21 VP5 FL C his	pOPINE	1485	55.9
14625	BTv21 VP5 FL C his	pOPINF	1493	57.0
14626	BTv26 VP5 FL C his	pOPINE	1596	59.9
14627	BTv26 VP5 FL C his	pOPINF	1604	60.9
14628	BTv26 VP5 FL C his	pOPINE	1536	57.8
14629	BTv26 VP5 FL C his	pOPINF	1544	58.9
14630	BTv26 VP5 FL C his	pOPINE	1476	55.7
14631	BTv26 VP5 FL C his	pOPINF	1484	56.8

3.2.3. Constructs expression and solubility test

All constructs listed in Table 3.1 were co-transfected with linearized, essential gene orf1629 knocked out baculoviral genomic DNA, this ensured only recombinant baculovirus could be generated without virus plaque purification. Small scale expression tests used insect sf9 cells in 6-well dishes to infect recombinant baculovirus. Infected cells were collected and subject to hypotonic treatment. Then, after centrifugation, samples were separated into supernatant and pellet fractions. Both fractions were analysed in SDS-PAGE gels to identify expression level and solubility. For example, the expression and solubility tests from BTV serotypes 3, 4, 8, 19, 21 and 26 are shown in Figure 3.4. From the expression test, 14 out of 36 constructs were expressed at a high level with clearly visible bands corresponding to expected size on SDS-PAGE. Among them, 2 are full length constructs, 6 are with 21 amino acids truncated and 6 with 41 amino acids truncated. However, the solubility tests show only 4 constructs are well soluble. They are BTV4-VP5- Δ 41-Nhis; BTV4-VP5- Δ 41-Chis; BTV8-VP5- Δ 41-Chis; and BTV26-VP5- Δ 41-Chis.

Further expression and solubility tests from BTV10 and BTV15 are shown in

Figure 3.5. The four truncated constructs of BTV15 show a better expression level than the others, among which the BTV15-VP5- Δ 43-Nhis is showing a clear protein band in the supernatant. Some of the others are also soluble but at lower levels. To further confirm the observed bands were actually from the targeted proteins, the Western blots with anti-penta his antibody were performed. As shown in

Figure 3.5 panel B, the observed bands at the expected size were indeed Western blot positive.

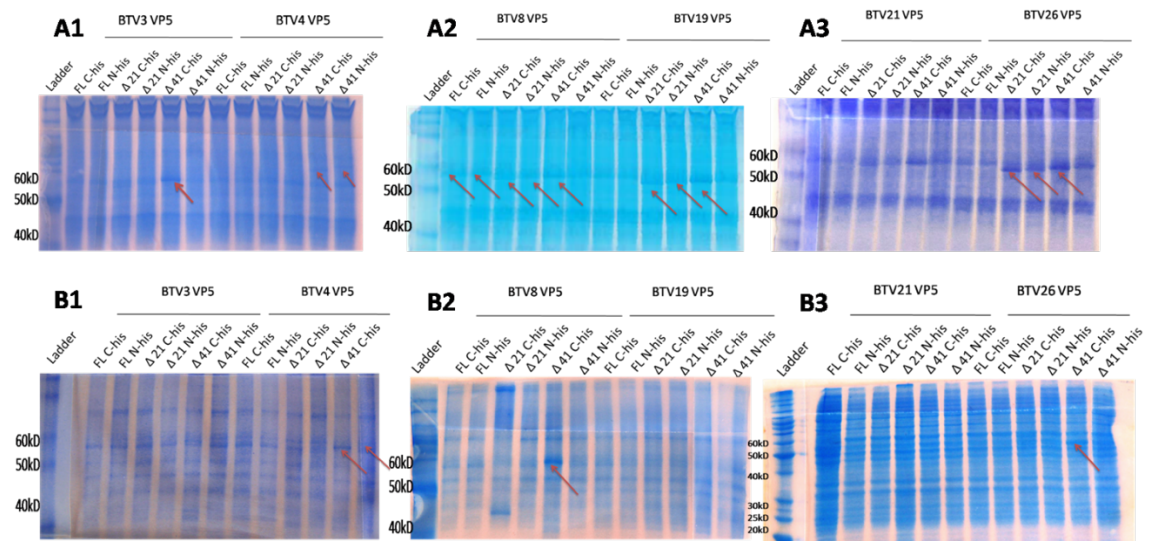


Figure 3.4 Expression and solubility test of BTV serotype 3, 4, 8, 19, 21 and 26 VP5 proteins by SDS-PAGE analysis of (A1-3) whole cells and (B1-3) supernatant. Arrows point to the constructs which show better expression or solubility than the other constructs

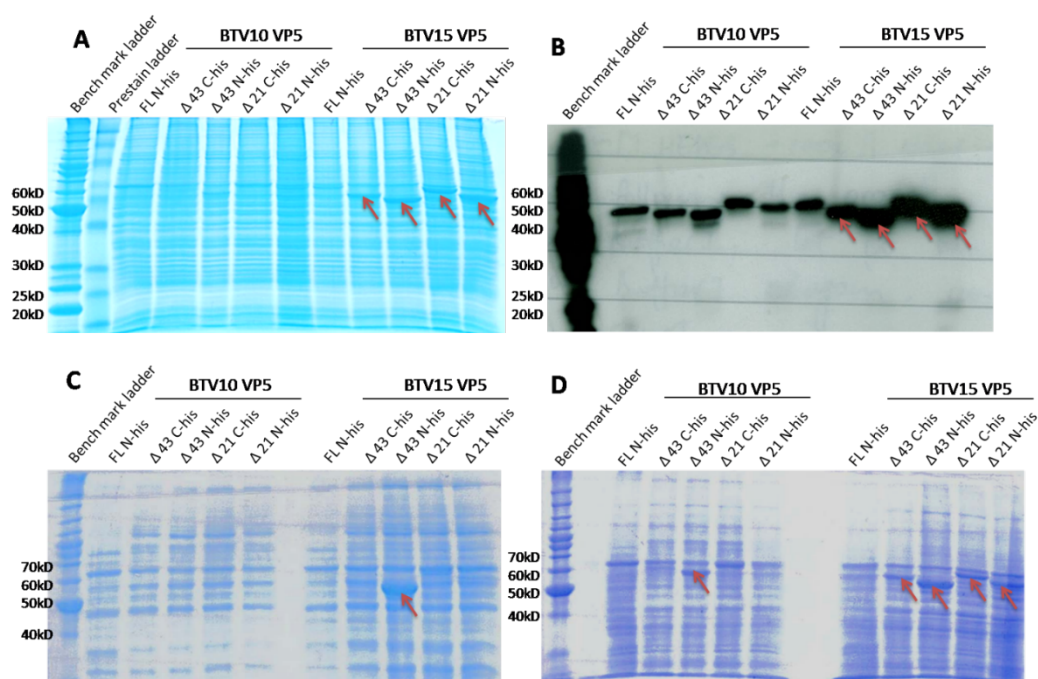


Figure 3.5 Expression and solubility tests of BTV10 and BTV15 VP5 proteins. (A) Whole cells samples analysed by SDS-PAGE. Arrows point to the four truncated constructs of BTV15 which show better expression than the other constructs. (B) Western blot experiment of whole cells. (C) Supernatant after centrifugation of the disrupted cells. (D) Pellet after centrifugation of the disrupted cells.

These trial expression experiments indicated that the N-terminal amphipathic α -helices (Figure 2.1) have a tendency to cause a cytotoxic effect when expressed in a recombinant baculovirus system. The full-length constructs were expressed poorly, while deleting one or both of the amphipathic helices can boost expression. Thus, BTV15 VP5 construct with these amphipathic helices removed was chosen for further large scale expression and protein purification.

3.2.4. Large scale protein expression and protein purification of BTV15 VP5

Virus harbouring the selected construct (BTV15-VP5- Δ 43-Nhis) was amplified and used to infect Sf9 cells in a few 2 Litre flasks with about 300 ml medium in each flask. The multiplicity of infection (moi) was about 10:1 for protein production. The infected cells were incubated at 27.5°C for 72 hours with gentle shaking (100 rpm). The cells were harvested by centrifugation. The purification procedures include cell breaking, centrifugation, affinity catching with cobalt beads and eluting with imidazole. The eluted proteins were further polished by size exclusion chromatography with a Superdex S200 column connected to a AKTA purifier machine. Example details of the protein purification are shown in Figure 3.6. The whole cell lysate and cell lysate supernatant fractions contain strong protein bands of expected size, whereas flow through or washing fractions contain little target protein bands. After washing with 10 mM imidazole, the fraction eluted with 250 mM imidazole showed high protein purity (Figure 3.6, top panel). The eluted fraction was subject to further polishing with Superdex S200 size exclusion chromatography. As shown in the Figure 3.6, bottom panel, after the sample was injected into a self-packed S200 10/300 column, the aggregates come out quicker, before 13 mls, while non-aggregated proteins come out in 16-18 mls. This procedure successfully removes the aggregated proteins and produces protein of crystallisation quality.

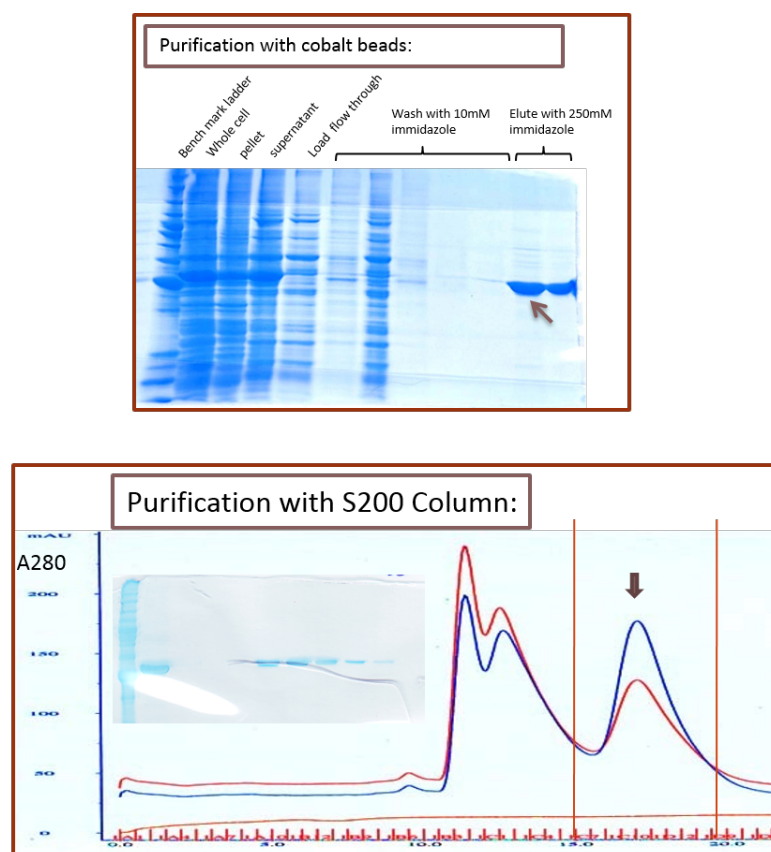


Figure 3.6 BTV15-VP5- Δ 43-Nhis protein expression with Sf9 cells and purification with a Cobalt beads affinity column and an S200 size exclusion column. Arrows point to the band and peak of the purified protein.

When this project was started, there was no structural information available from any BTV VP5. In order to obtain experimental phasing information, a seleno-methionine (Se-Met) protein labelling approach was used. There are 16 methionine out of total number of 504 residues in the construct of BTV15-VP5- Δ 43-Nhis. The Se-Met labelled protein was expressed with a special Methionine free SF900-II medium supplemented with seleno-methionine. The purification procedures were similar to those used for the native protein as described above. An example of an SDS PAGE result and size exclusion chromatography profile was shown in Figure 3.7. The protein yields are about 2 mg per litre of culture.

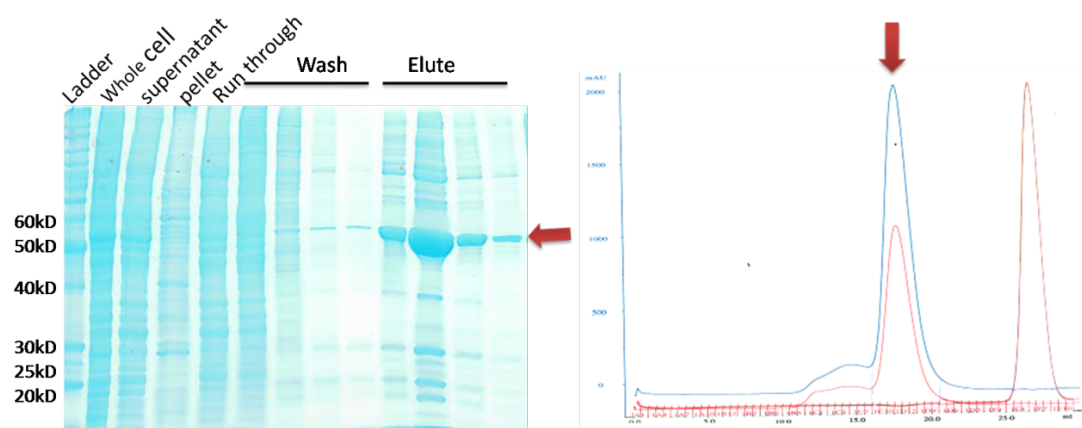


Figure 3.7 Se-Met labelled BTV15-VP5- Δ 43-Nhis protein expression with Sf9 cells and purification with a S200 size exclusion column. Arrows point to the band and peak of the protein (on left are blue protein bands and on right are size exclusion chromatography peaks. Red and blue traces represent 260 and 280 nm absorbance respectively).

3.2.5. Protein characterizations by MultiAngle Light Scattering (MALS)

Although the superdex S200 size exclusion chromatography profile can be used for estimate the oligomeric status and molecular mass in solution, MultiAngle Light Scattering (MALS) can be used to precisely determine the absolute molar mass and the average size of particles in solution by detecting how they scatter light. The purified VP5 protein was subjected to Multiangle light scattering analysis using an analytical Superdex S200 10/30 column with online static light scattering, differential refractive index and Agilent 1200 UV detectors. The results show a peak mass of 114.1 kDa for the VP5 construct (Figure 3.8A). The instantaneous concentration of the protein at the top of the peak was 0.05mg/ml. The calculated molecular mass for the monomeric VP5 was 56.6 kDa. The observed trend of MALS mass follows the concentration of the protein which suggests that the protein is in equilibrium and it has not reached its maximum mass. The curvature of molar mass observed along with the

change of the protein concentration suggested that the oligomeric state of the VP5 protein may be concentration dependent. Increasing the concentration of the protein injected may allow larger oligomers to be observed. The elution of the VP5 protein from the column is shown in the chromatogram (Figure 3.8B).

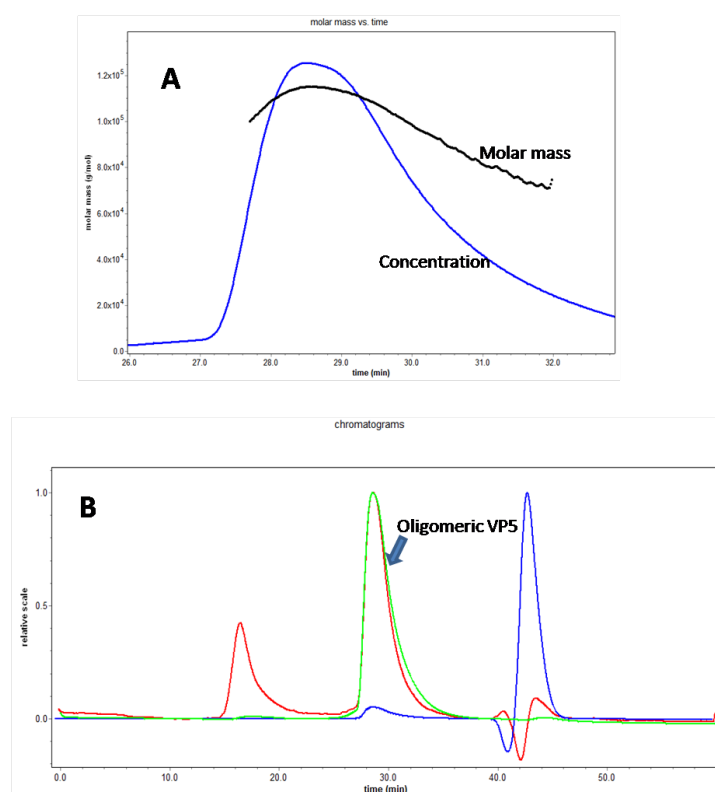


Figure 3.8 MALS of the protein from construct BTV15-VP5- Δ 43-Nhis. (A) Molar mass of protein determined under an injection concentration of 1mg/ml. (B) chromatogram of the MALS run (red: rayleigh ratio; blue: refractive index; green: UV light).

3.3. Protein Crystallization

Judging from size exclusion chromatography profiles, MALS results and SDS-PAGE gel, it suggests that the protein quality is good enough for protein crystallization trials. The protein samples were concentrated to 5-10 mg/ml. In order to determine the optimal protein concentration for the crystallization trial, a Hampton PCT test was performed. After mixing the protein with 4 different concentrations of salt or PEG based precipitants, it was determined a concentration of about 8 mg/ml would be a good starting point.

Crystallization trials were performed in a sitting drop vapour diffusion format using 100 nl protein solution adding into a 100nl reservoir solution set up in 96-well Greiner plates with the Cartesian Technologies robots (Walter et al. 2005). Dispensed plates were kept in a 20°C imager. A total of 14 blocks of different conditions (each block contains 96 conditions) were used for screening for protein crystals. Crystals appeared in both low (pH 6.0) and high (pH 9.0) pH conditions. For the high pH condition, the crystals were formed in an optimal condition of 0.09 M Sodium Formate, 6.6% PGA-LM and 0.1 M Tris (pH 9.0). The crystals appeared in a week. The rhombic-shaped crystals grew to a size of approximately 100×50 microns (Figure 3.9). For the low pH condition, the crystals appeared in two weeks in the condition of 1.0 M Sodium Chloride and 0.1 M Tris (pH 6.0). An example of a monoclinic-shaped crystal appeared to be approximately 200×150 microns is shown in Figure 3.11 left panel. The crystals were fished out with a litho loop and soaked in mother liquor supplemented with 25% (v/v) ethylene glycol before freezing in liquid nitrogen.

3.4. Data collection and processing

X-ray data collections were carried out at Diamond synchrotron radiation Light Source. The data from BTV15 VP5 Se-Met crystals were collected on beamline I03 with a Pilatus3 6M detector. Prior to data collection, a fluorescence scan was carried out. The X-ray fluorescence spectra is shown in Figure 3.10A. The CHOOCH chart in Figure 3.10B indicated a peak at energy 12656.99 eV which represents the signal from selenium. This confirms the existence of the Se-Met in the crystal. Data were collected slightly away from the absorption edge at 0.97918 Å corresponding to an energy of 12662.0 eV. Five datasets of 180 degrees were collected with a 50µm ×20µm beam size. The oscillation was set at 0.1°. A total of 1800 images were taken for each of the datasets with 0.02 seconds exposure per image. The best diffracted data set showed visible reflections up to ~2.8 Å. The crystal belongs to space group P1, unit cell 91.4 Å, 142.2 Å, 142.15 Å; α 119.5°, β 96.1°, γ 95.8°. The data were processed by HKL2000 (Zbyszek & Wladek 1997) and the statistics summarised in Table 3.2.

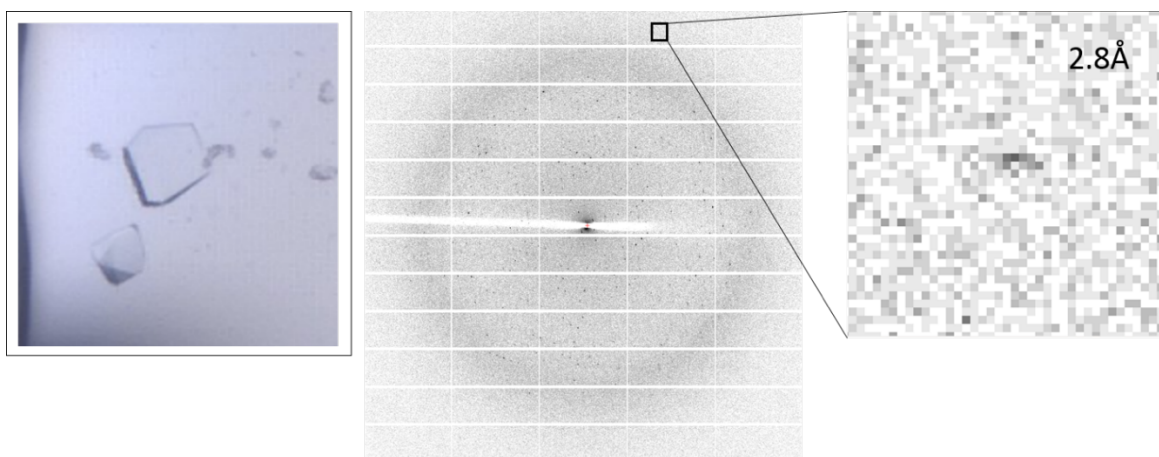


Figure 3.9 BTV15 VP5 protein crystals and collected data images. The left panel shows the crystals. The middle panel shows a typical image from the data collection and the right panel shows a close up area at resolution around 2.8 Å.

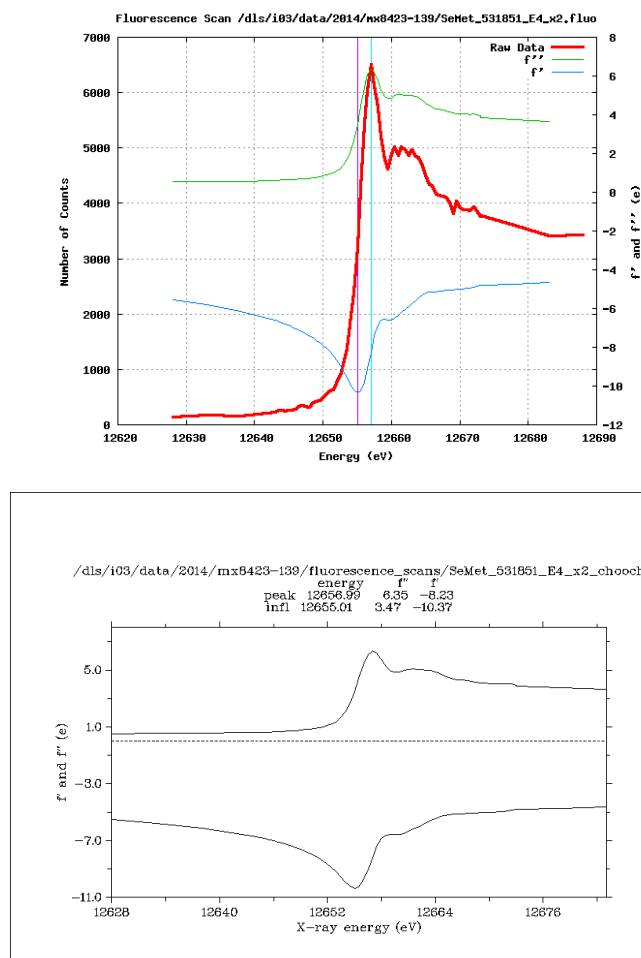


Figure 3.10 X-ray fluorescence scan plot from the Se-Met labelled VP5 crystal. The top panel is the raw scan plot and is coloured red, f' is coloured blue and f'' is coloured green. The vertical purple line indicates the peak of f' while the cyan vertical line indicates the peak of f'' . The lower panel is from CHOOCH (Evans & Djinovic 2003). The best X-ray energy for maximum anomalous is calculated and shown as the peak wavelength in the bottom panel.

The data for BTV15 VP5 native crystals (pH6) were collected on beamline I24 with a Pilatus3 6M detector. Data were collected from several crystals, with 360 degrees collected from each, with a $20\mu\text{m} \times 20\mu\text{m}$ beam size and a rotation of 0.05° per image. Each of the three data sets comprised 3600 images with 0.02 seconds of exposure time per image. The best data set showed visible reflections up to $\sim 2 \text{ \AA}$. Auto-indexing and processing was performed by Xia2 3dii (Winter 2010) upon data collection. The crystal belongs to the H3 space group. After the unit cell sizes were examined, the two data sets with similar unit cell size were merged together by Xia2 3dii. The final working native dataset was at a resolution of $2.19 \text{ \AA}$ with a space group of H3 and unit cell of $76.9 \text{ \AA}$, $76.9 \text{ \AA}$, $163.1 \text{ \AA}$, 90° , 90° , 120° . These statistics are summarised in Table 3.2.

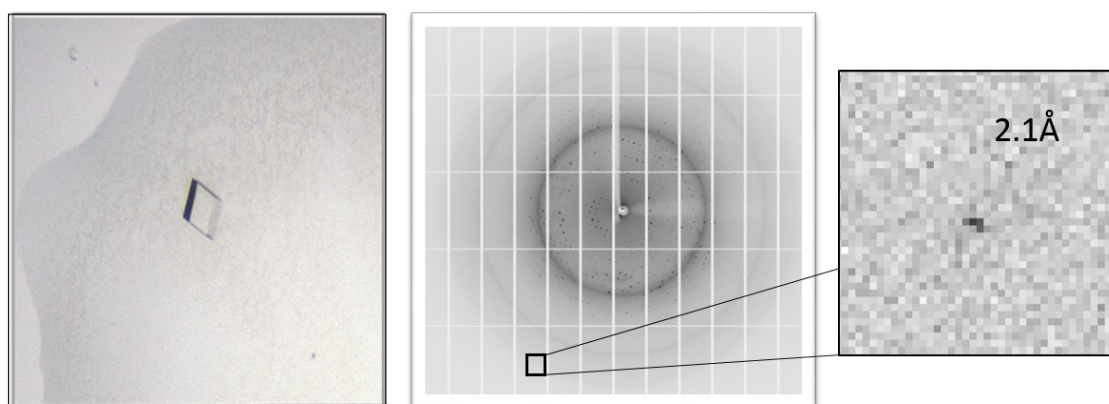


Figure 3.11 Crystal morphology and diffraction pattern of the crystal formed from the low pH condition. Left panel shows a crystal in a crystallization drop. Middle panel shows a typical image of diffraction and the right panel shows a close up area at $2.1 \text{ \AA}$.

Table 3.2 Data collection and refinement statistics for VP5 crystals at two different pHs.

Data collection	High pH (pH9)	Low pH (pH6)
Wavelength (Å)	0.97918	0.96864
Space group	P1	H3
Cell dimensions:		
a, b, c (Å)	91.4, 142.2, 142.2	76.9, 76.9, 163.1;
α, β, γ (°)	119.5, 96.1, 95.8	90, 90, 120
Resolution (Å)	46.6-2.80 (2.90-2.80)*	51.8-2.19 (2.25-2.19)
Total reflections	147513 (15781)	18627 (12139)
Unique reflections	290549 (30086)	18479 (1322)
R_{merge}	0.12 (0.67)	0.07 (---)
CC₅₀	0.99 (0.64)	0.99 (0.65)
I / σI	11.0 (1.7)	13.6 (2.2)
Completeness	98.9 (98.4)	100 (100)
Redundancy	3.5 (3.5)	9.9 (9.2)
Refinement		
Resolution (Å)	46.6-2.80	51.6-2.19
No. reflections	147486	36944
No. atoms		
Protein	30625	1972
Water/other	---	12
R_{work} / R_{free}	0.22/0.26	0.20/0.21
Average B-factor (Å²)		
Protein	70	83
Water/other		72
R.m.s. deviations		
Bond lengths (Å)	0.003	0.003
Bond angle (°)	0.6	0.5
Ramachandran plot		
Favoured	96.39%	98.1%
Outliers	0.44%	0.47%

*Numbers in brackets refer to the highest resolution shell of data.

3.5. The structure of BTV15 VP5 at high pH

Data with Se-Met labelled protein were collected as at the time of the project progressing, there wasn't any similar structure available. In order to obtain phasing information, the data were collected at a wavelength of 0.97918 Å (corresponding to an energy of 12662.0eV) to maximise the selenium anomalous signal. The anomalous signal could be detected in the data set when using the programmes of HKL2map and SHELX C/D/E and the Se-Met positions were located. With this phase information, it should be possible to build the structure. As the space group is P1 with 9 monomers in a crystallographic unit cell, building the full model would be a challenging task at a resolution near 3 Å and it would take a long time. During the process of model building, the VP5 cryo-EM structure at a similar pH (pH 8.8) became available (Zhang et al. 2016). This available structure was used as a search model to perform molecular replacement and this facilitated model building. The structure was refined with repetitive cycles of Refmac5 (CCP4 package) and rebuilt using Coot before finally being refined with Phenix, with final R_{work} 22%, R_{free} 26%, and good geometry. The torsion angle was chosen for Non-Crystallographic Symmetry (NCS) restraint. Individual B factors are refined in addition to the TLS parameters. The refinement statistics are shown in Table 3.2. One of the potential drawbacks for the molecular replacement approach is the possible introduction of model bias. To ascertain our model is correct and correlates well with the Se-Met phasing information obtained, an anomalous difference map was generated, as shown in the Figure 3.12 and Figure 3.13.

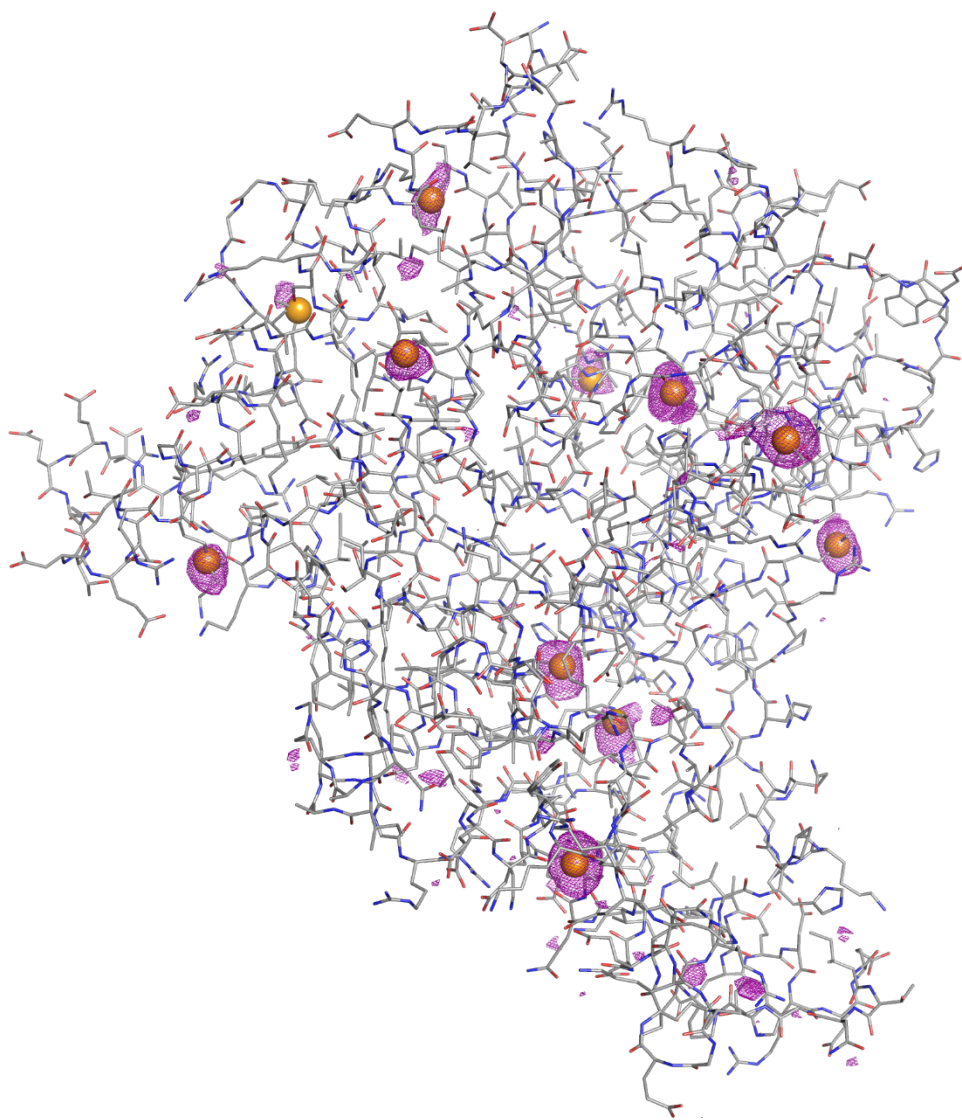


Figure 3.12 Anomalous density map for Se-Met in BTV15 VP5. The anomalous density from Se-Met is shown as purple mesh with contour level at 3 σ . The difference map is created by Anode and the picture is drawn using Pymol. The protein is drawn as sticks with the Met-SD (Mse-Se) highlighted as balls.

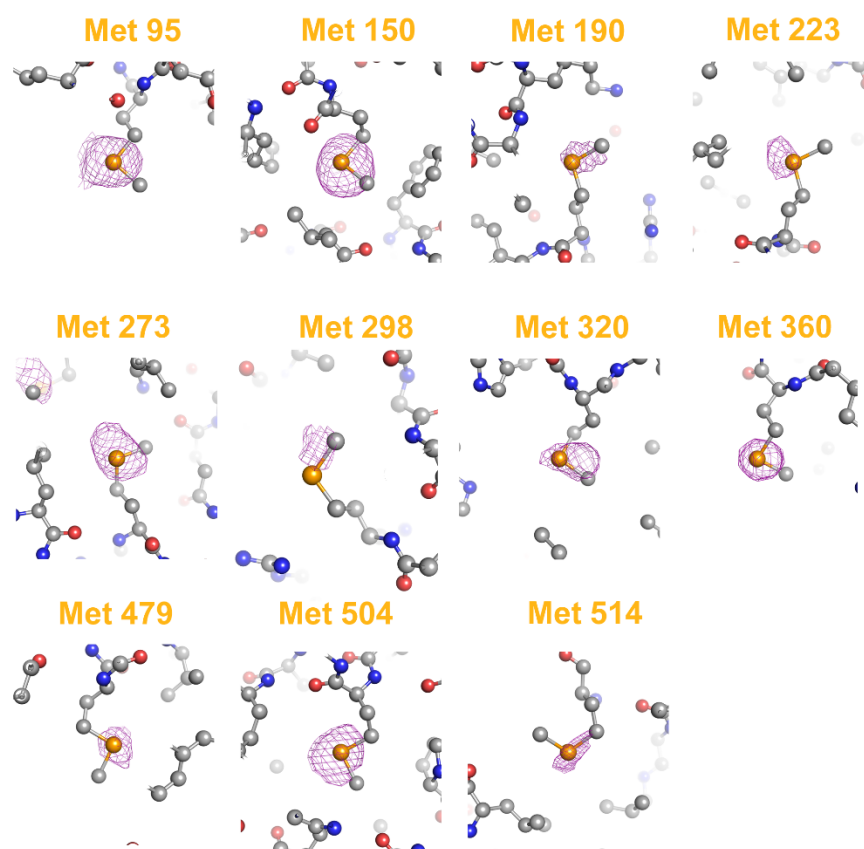


Figure 3.13 Detailed anomalous density map with labelled Se-Met numbers in BTV15 VP5. The protein is drawn as ball and sticks. All the Met residues are shown as brown balls and the density map shown as purple mesh, contoured at 4σ (except M298 which is contoured at 3σ).

3.5.1. The overall structure

The BTV15 VP5 construct omitted 43 N-terminal amino acids and replaced them with a 6×his tag. Thus, this construct omitted the first beta strand and first alpha helix (as named in BTV-1 cryo EM structure, PDB code: 3J9E). The resultant structure resolves 428 out of 491 residues within the constructs of BTV-15 VP5. The N-terminal his tag, the second amphipathic helix ($\alpha 2$), the loop region between $\alpha 15$ and $\beta 8$ and the final 13 C-terminal residues were too disordered to be seen clearly in our structure. The full BTV-15 VP5 amino acid sequence and the corresponding secondary structure is shown in the Figure 3.14.

The resolved structure starts from the $\beta 2$, and finishes with $\beta 9$. These two N-terminal and C-terminal β strands come together in an anti-parallel orientation and are further strengthened by a third β strand ($\beta 3$). The cartoon presentation of the monomer is shown in Figure 3.15 with front and back views. Cartoons with surface representation depicting front and top views are shown in Figure 3.16. The monomer shows a “palm” shape, with the top half termed the unfurling domain(K70-F355) sitting on the bottom half which has been termed the anchoring domain (I356-A527) as described in the BTV-1 cryo-EM structure (Zhang et al. 2016). These two domains are depicted in the Figure 3.17, with unfurling domain coloured blue and anchoring domain coloured green. The unfurling domain is rich in alpha helices, from $\alpha 3$ to $\alpha 14$. The $\alpha 3$ is also described as the “stem” helix. It is in cross support of two horizontal long helices, $\alpha 6$ and $\alpha 12$ which are termed “beam” helices. The bottom part of anchoring domain, located toward the C-terminal region, contains 4 antiparallel β strands ($\beta 4$ - $\beta 7$) sandwiched between $\alpha 15$ and $\alpha 16$ -18 helices. The 4 β strands are meandering in shape and are termed “ β meander motif”. The motif is further augmented by another c-terminal β strand ($\beta 8$).

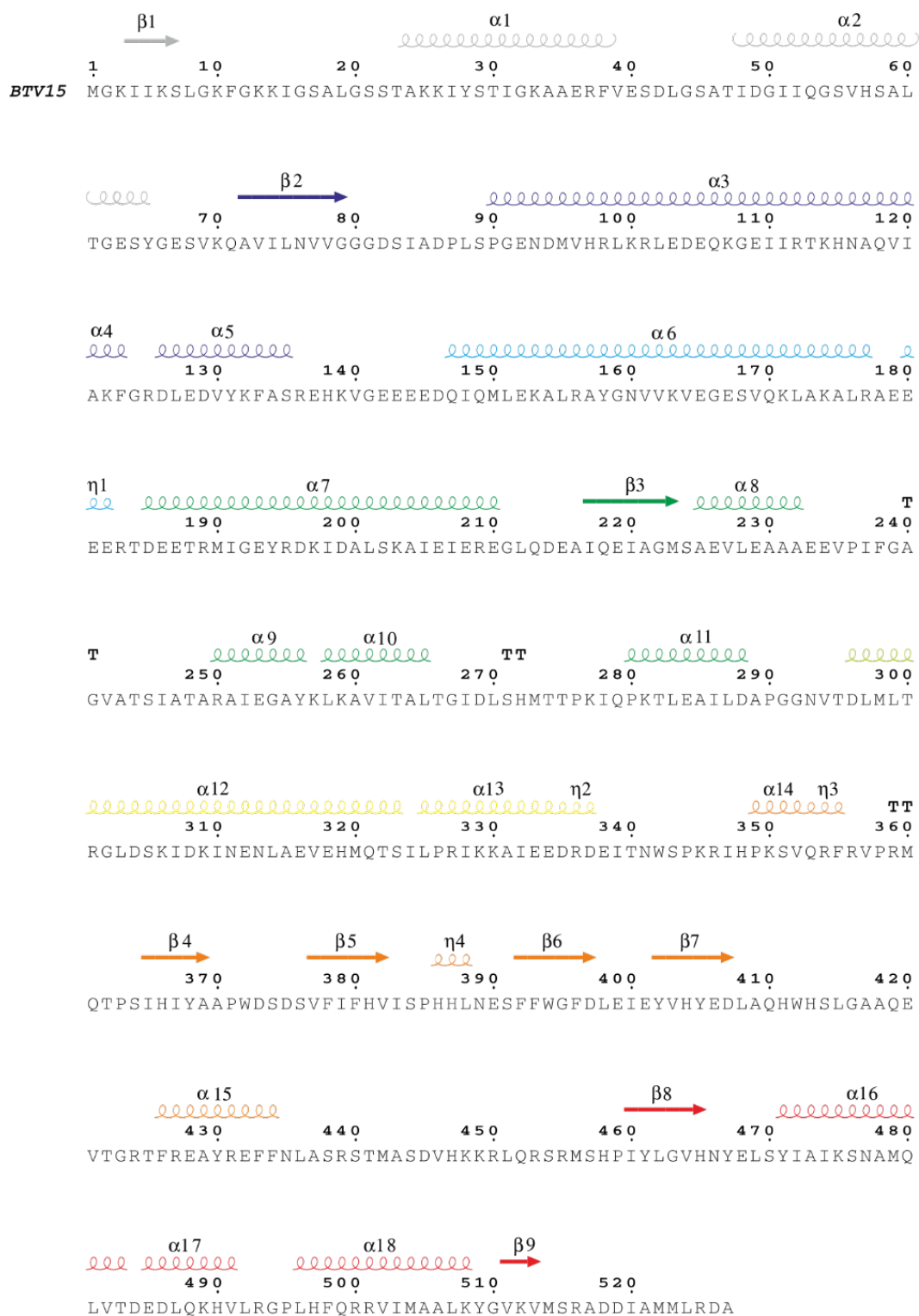


Figure 3.14 BTV-15 VP5 amino acid sequence annotation. The secondary structure is shown above the amino acids sequence, coloured from N-terminal blue to C-terminal red. The N-terminal 43 amino acids are not included in the construct and the predicted secondary structure is shown in grey colour. “ η ” indicates 3/10 helices.

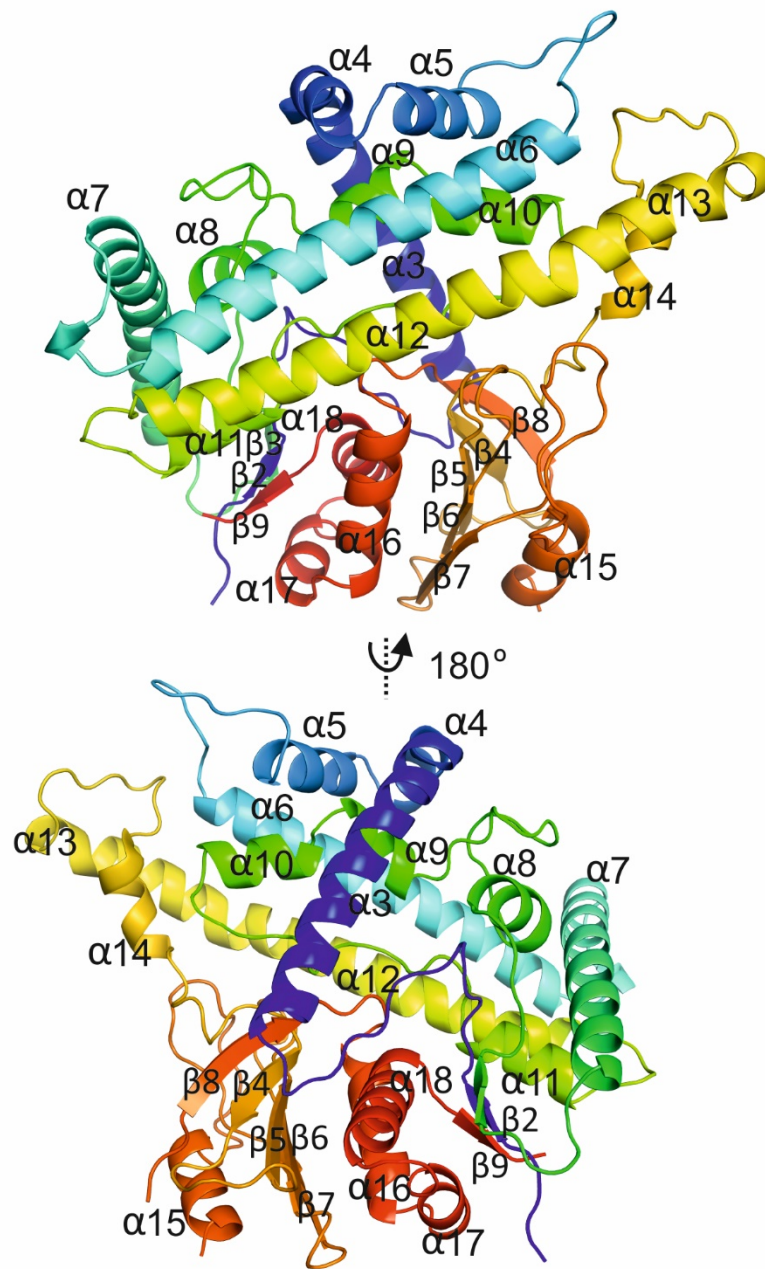


Figure 3.15 Cartoon representation of BTV-15 high pH VP5 structure. The bottom panel is the back view of the top panel, the cartoon is coloured in rainbow, from N-terminal blue to C-terminal red. The secondary structures α helices and β strand are numbered consecutively according to the full-length VP5.

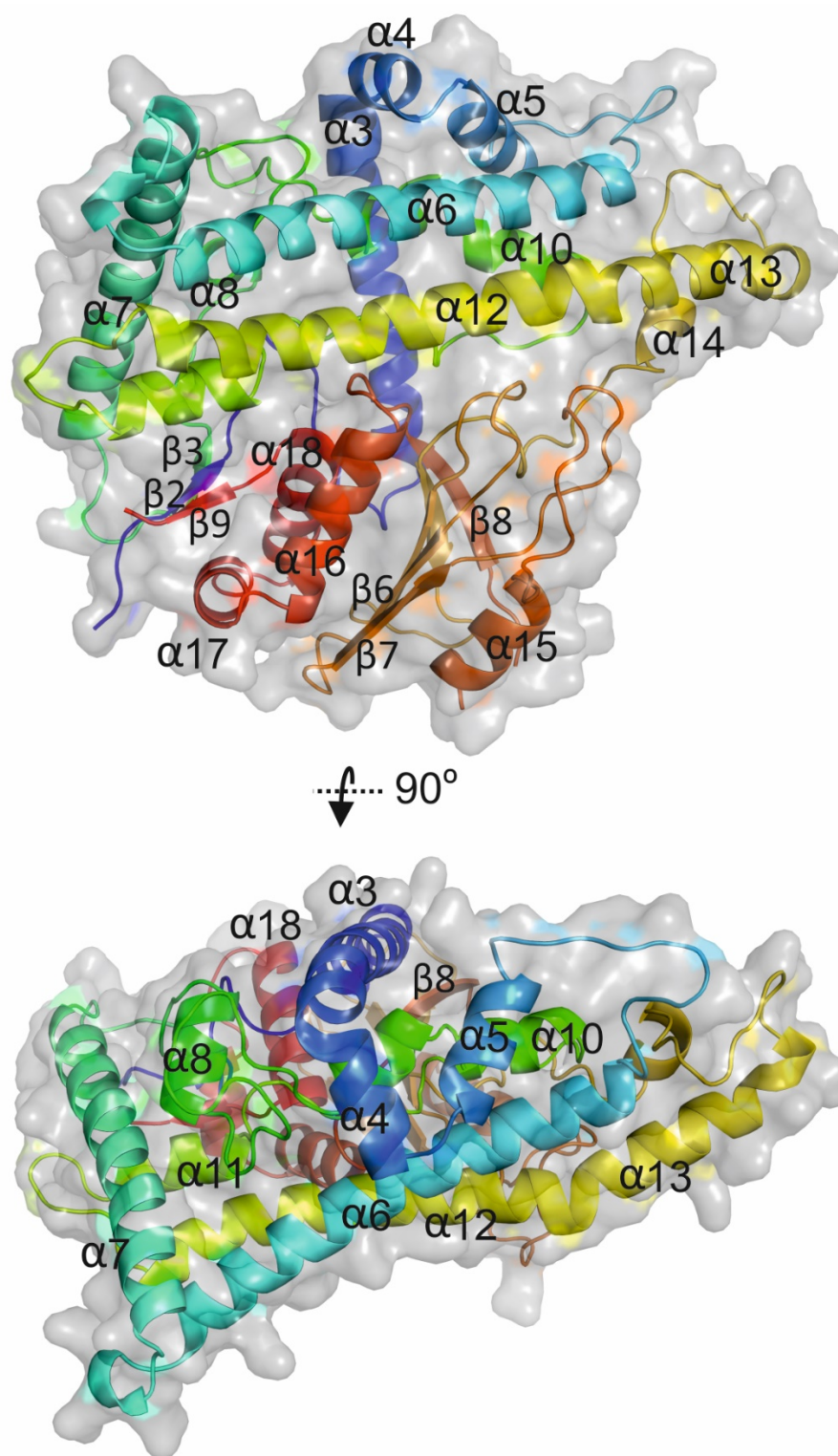


Figure 3.16 Cartoon and surface view of BTV-15 high pH VP5 structure. The bottom panel is the top view of the above panel, the cartoon is coloured in rainbow, from N-terminal blue to C-terminal red. The surface is coloured grey, with 50% transparency.

3.5.2. Comparison of crystal structure and cryo-EM structures

Comparing the recently reported BTV-1 cryo-EM structure at similar pH, the overall structures align well, as shown in the Figure 3.17. The Structure Homology Programme (SHP) was employed to calculate the similarity (Stuart et al. 1979). The numbers of accepted residues from crystal and cryo-EM structures were 428 and 520 respectively. There were 225 sequence identities out of 408 equivalences (55.1%). The average distance is 0.91 Å. The RMS distance is 1.06 Å. The distance of corresponding C α s from both the crystal structure and cryo-EM structure is plotted in Figure 3.18. The majority of residues are at distance around 1 Å, with a few of them show distance up to 3.5 Å.

As shown in Figure 3.17, the all alpha helices and beta strands are well aligned. Some loops, especially the loop between α 5 and α 6, and the loop between α 8 and α 9 within the unfurling domain show large changes. It is noteworthy that the cryo-EM structure's Dagger domain is flexible (Zhang et al. 2016) with a long linker. If the full dagger domain was included in our constructs, it might make the protein uncrystallisable.

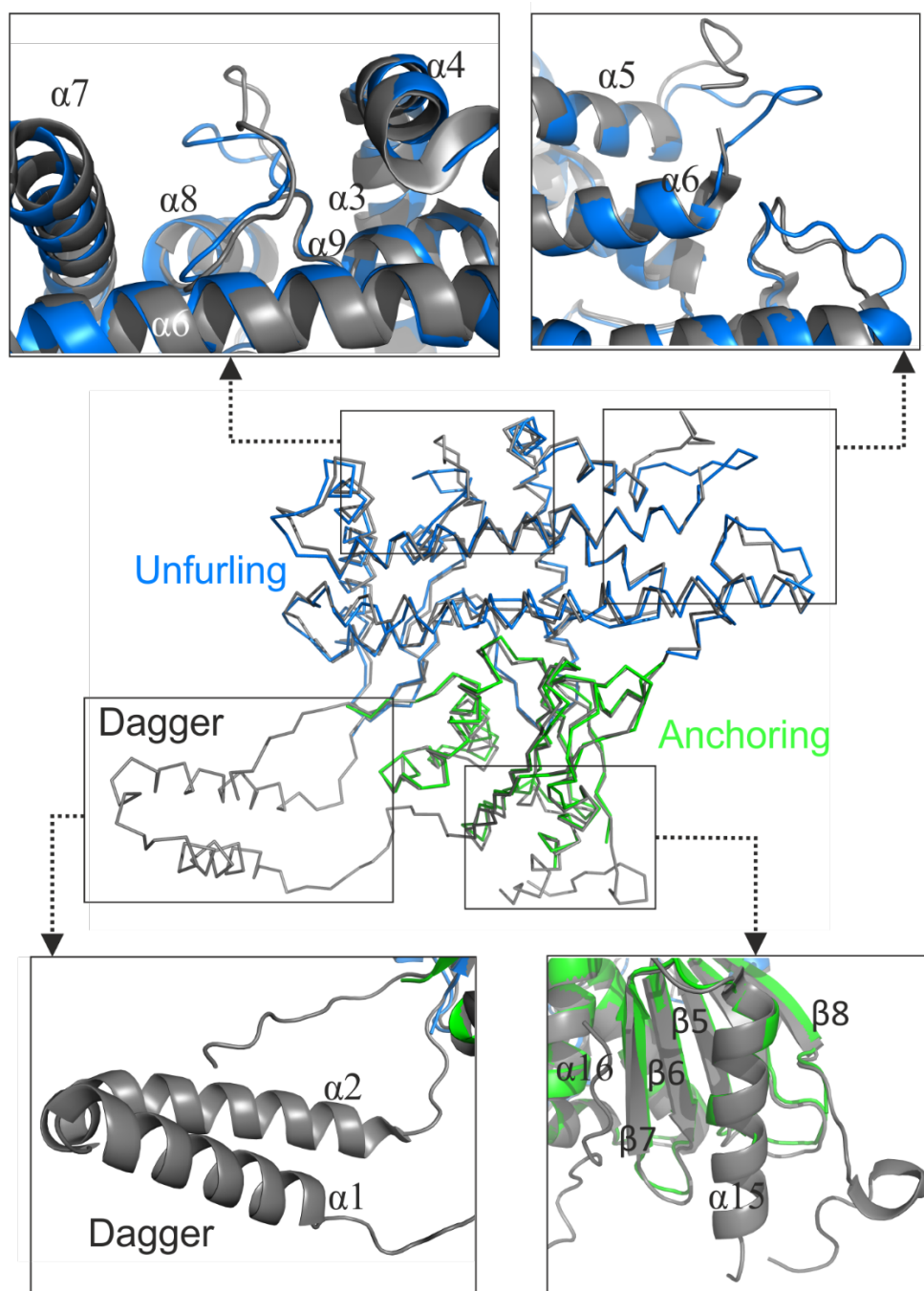


Figure 3.17 Alignment of VP5 crystal structure and cryo-EM structure both at high pH. The cryo-EM structure is coloured grey, including the dagger domain which is only exist in cryo-EM structure. In the crystal structure, the unfurling domain is coloured blue and the anchoring domain is coloured green. The close up areas are indicated in corresponding boxes.

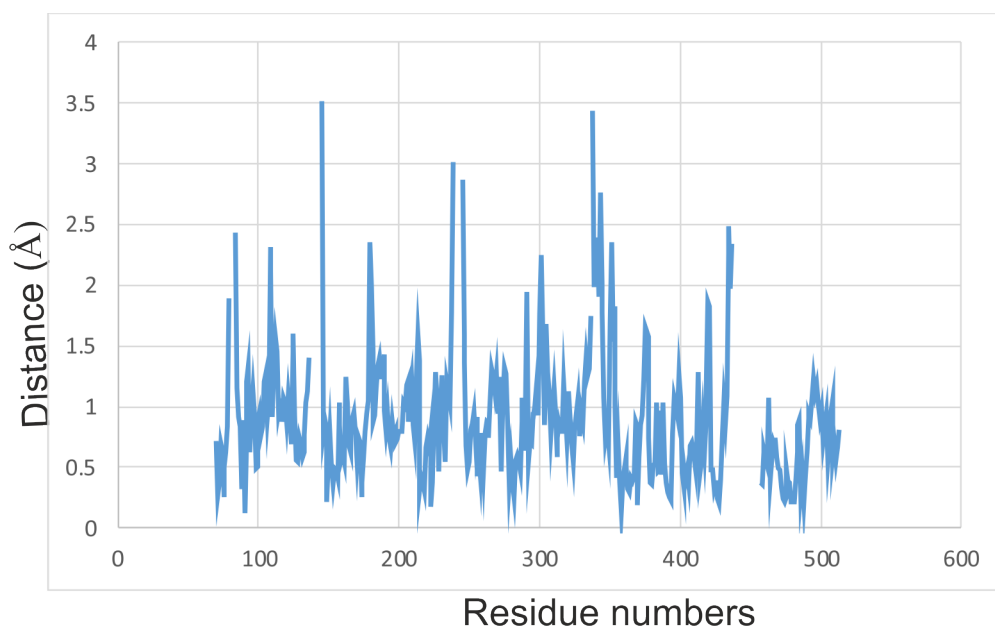


Figure 3.18 Plot of distance between corresponding C α s of the crystal structure and cryo-EM structure (PDB code: 3J9E).

3.5.3. VP5 is a trimer in structure

The crystal belongs to the P1 space group, packed with 9 monomers in a unit cell. Within one unit cell, there are three trimers (Figure 3.19A). A biological trimer is shown in Figure 3.19 (B and C). In order to clearly show the organization among monomers, one monomer is drawn as a cartoon, another as a surface and the other as a mesh.

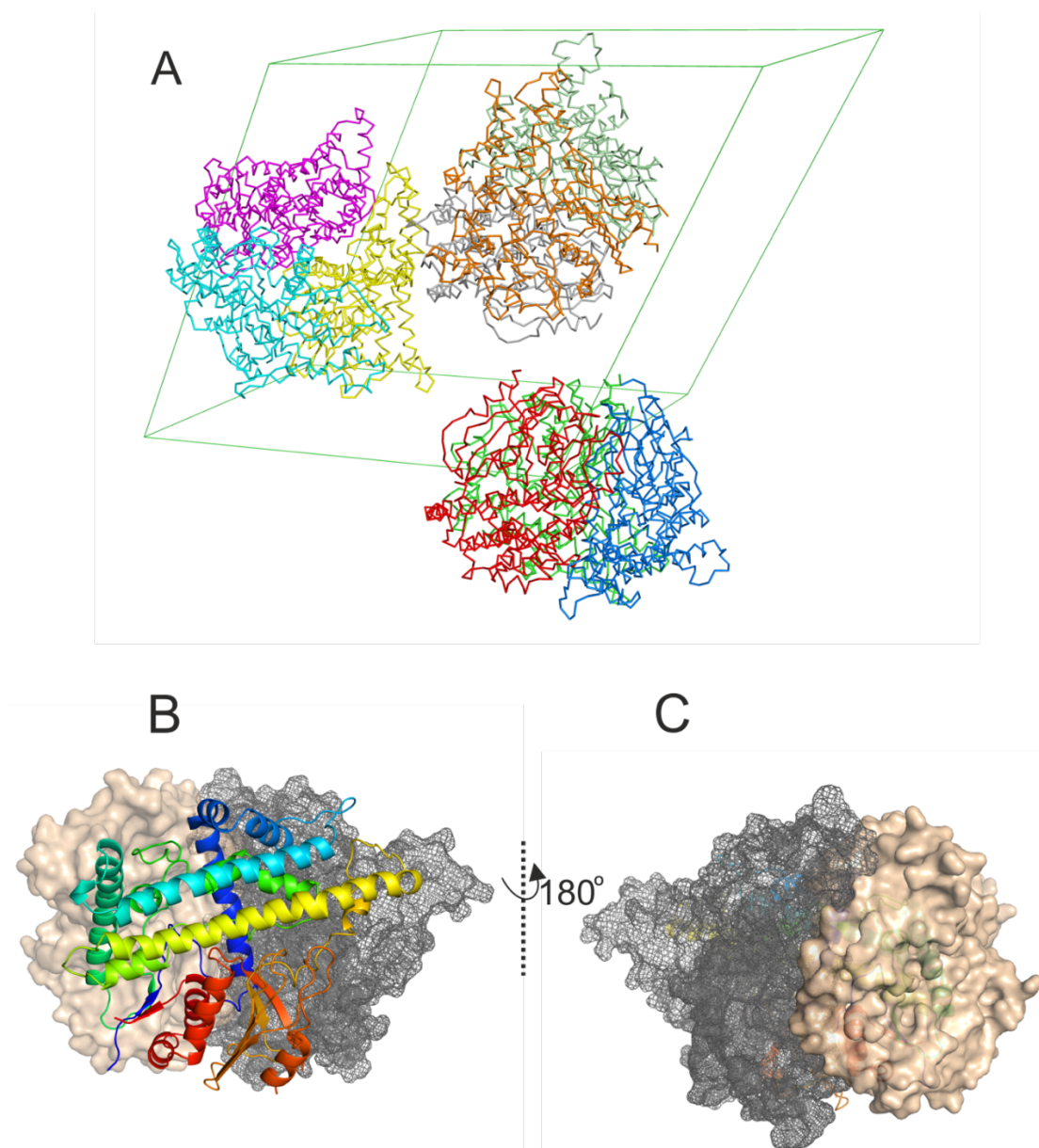


Figure 3.19 BTV VP5 crystal component in a crystallographic unit cell and trimer representation. (A) shows 9 monomers of VP5 in a unit cell organised as 3 trimers. The monomers are coloured differently. (B) shows a trimer; one monomer is present as cartoon, coloured rainbow; another is shown as beige coloured surface; The other monomer is shown as grey mesh. (C) is the back view of (B).

3.5.4. BTV VP5 shows a conserved surface for the subunit interfaces in the trimer, but more variable outer surface

The overall BTV VP5 sequence in all serotypes were initially aligned as shown in Figure 3.1. Once the crystal structure had been obtained, it became possible to see how the surface of the molecule is conserved. The consurf server (<http://consurf.tau.ac.il>) was used to plot the conserved surface among the 8 groups of serotypes of BTV strains. As shown in Figure 3.20, the outer surface (left two panel in the Figure 3.20) of BTV VP5 varies a lot; the majority of the surface is coloured cyan, indicating variation. Only small areas are coloured purple, showing conservation. However, the trimer contacts on the inner face show much more conserved areas. This suggests that even though the overall sequence identity is relatively high among different serotype groups (Figure 3.1), the exposed outside surface of the VP5 varied a lot among them. This may also suggest that antibodies generated from different serotypes may show limited cross-reactivity.

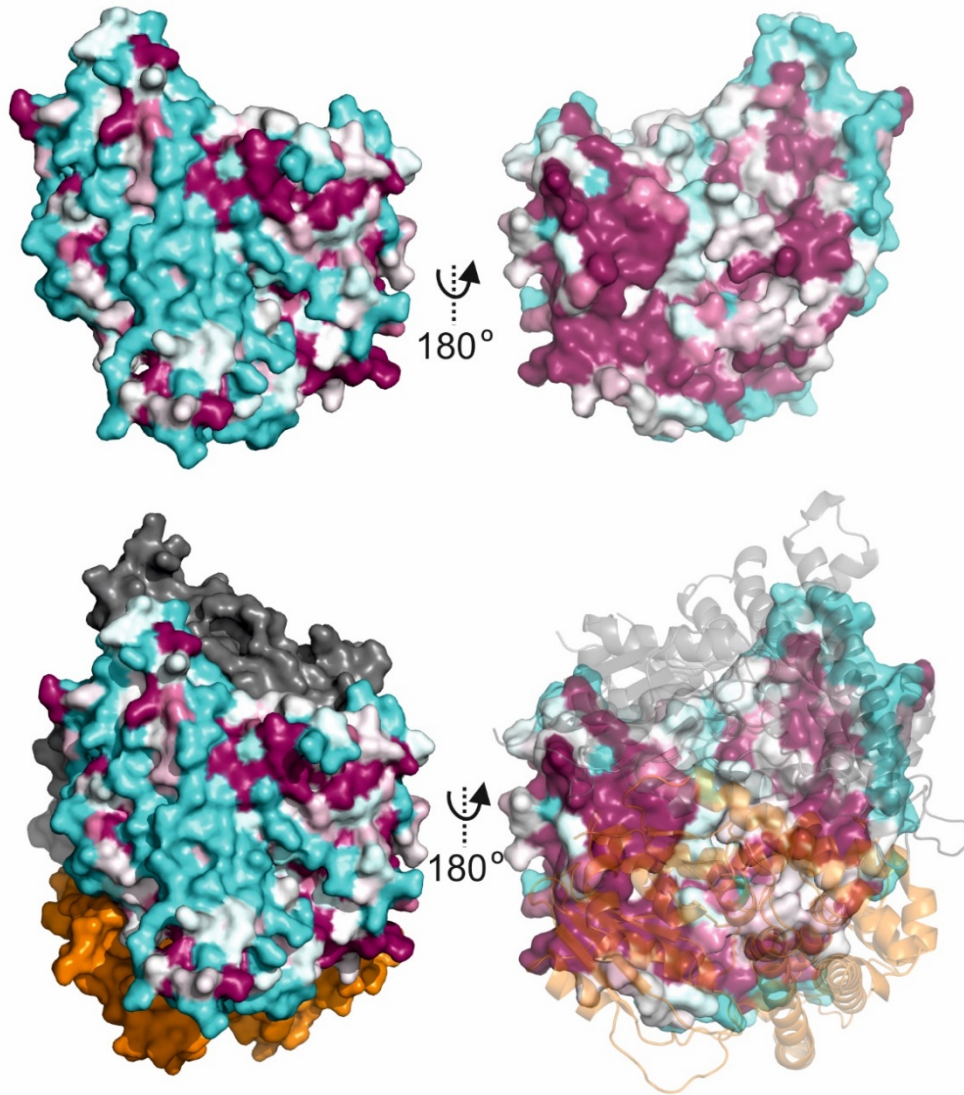


Figure 3.20 Surface conservation among different serotypes of BTV VP5 proteins. The purple colour represents conserved residues, cyan indicates non-conserved residues and white indicates the residues with moderate conservation. The top panel shows the monomer and the bottom panel shows the trimer. The bottom left panel shows a trimer, with one coloured same as the monomer above, another coloured grey and the other coloured brown yellow. The bottom right panel used half transparent cartoon to indicate the other two monomers covering the one showing conserved inner surface.

3.6. Structure of VP5 at low pH

Viruses can sense a cellular pH environment and change conformation in order to release the genetic contents for starting replication. The pH sensing and conformational change mechanism for non-enveloped viruses, like Bluetongue virus, is still poorly understood at the atomic level. After obtaining the high pH structure as described above, a low pH structure is wanted.

3.6.1. Model building for low pH VP5 structure

The crystal obtained in low pH (pH6) was grown in a simple salt (NaCl) condition. The data were collected at Diamond I24 beamline. The data set has a much smaller unit cell size than the high pH data set. The space group belongs to highly symmetric H3. The crystal diffracts to a high resolution limit of 2.2 Å. When the high pH structure was used as search model to perform molecular replacement, however no solution could be obtained by Phaser. When calculating the Matthews coefficient (Matthews Probabilities server <http://www.ruppweb.org/mattprob/default.html>) with the unit cell and expected amino acid length from the expression construct, the unit cell volume is 834638.4 Å³ and the asymmetric unit volume is 92737.6 Å³. The volume of protein (Matthews co-efficiency) is 1.69 Å³/Dalton. The resulting Matthews probabilities plot is shown in Figure 3.21. The estimated solvent content is 27.23%.

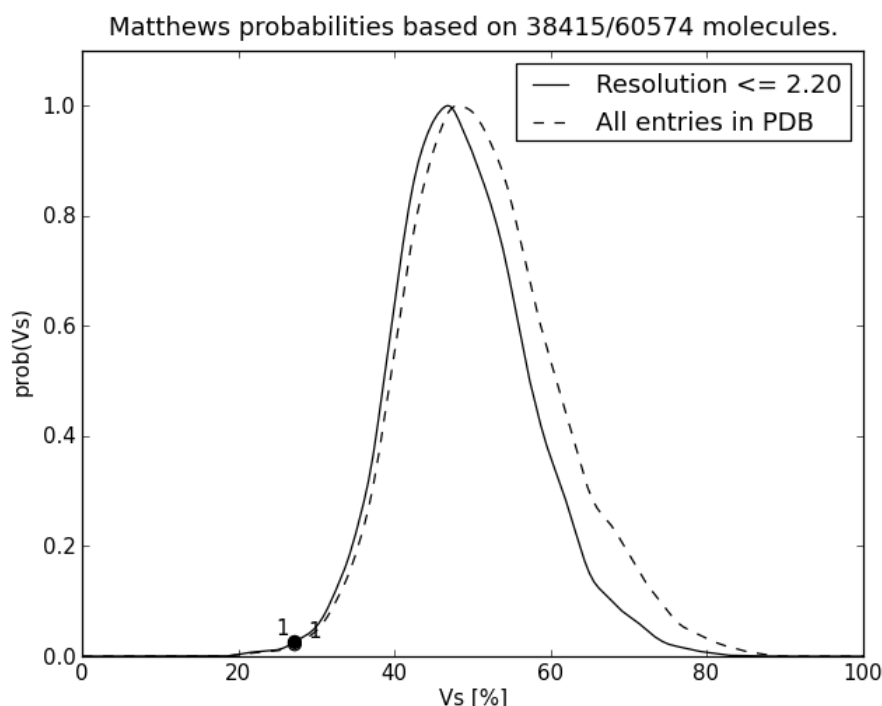


Figure 3.21 The Matthews probability plot for low pH crystal unit cell with all amino acids from the construct. The black dot indicates the possible number of monomers in the asymmetric unit. The plot indicates solvent volume distribution for overall pdb entries (dash line) and entries with resolution equal or less than 2.2 Å.

When a crystal diffracts to this resolution (2.2 Å), the most probable solvent content should be around 50% (Figure 3.21). So it seemed either some parts become disordered or even the crystal didn't contain the full molecule. Then the partial structures derived from the high pH VP5 were tested for searching a structural solution. When the unfurling domain was used as a search model, there was still no solution. Finally, when the anchoring domain was used, a structural solution was revealed. The anchoring domain fitted nicely into density as shown in

Figure 3.23. However, there was additional density, as indicated in red box of Figure 3.22. The extra density seemed to resemble a helix.

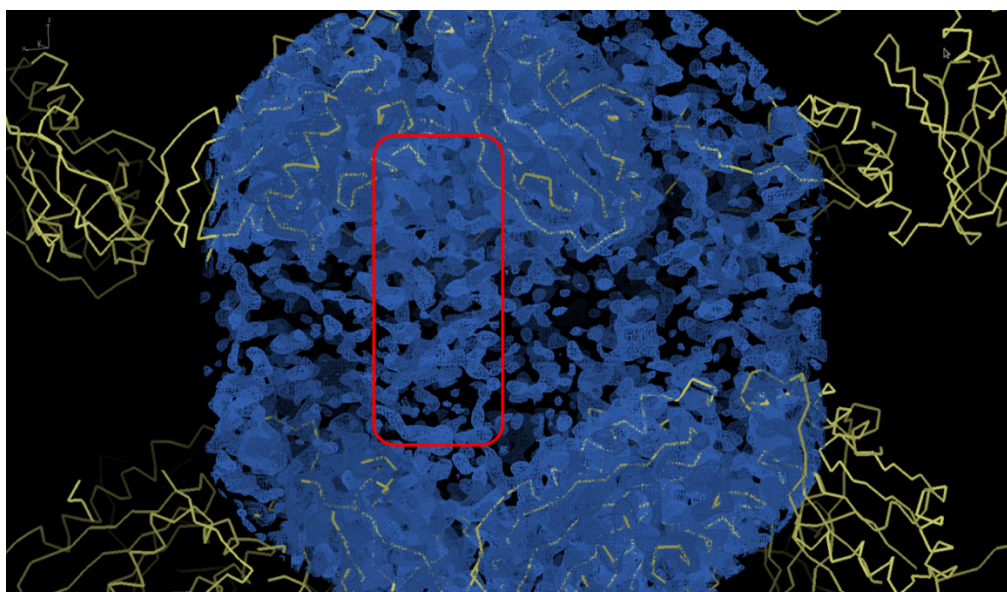


Figure 3.22 Electron density map from low pH crystal data when placed VP5 anchoring domain for phasing. The density map is shown as blue mesh, and anchoring domain C α is shown as yellow lines. The red square line showing the extra density apart from anchoring domain.

The extra helical density map prompts us to think there is some unfurling domain helical amino acids packed in the crystal. However, at the initially stage, it is not clear what part of unfurling domain sequence remain ordered in the structure. A poly alanine sequence was put into the helical density manually using Coot and performed refinement. This approach improved the phases, and thus the map density improved. As shown in

Figure 3.23, the helical electron density became much clearer. Figure 3.24 shows a typical area of electron density in Coot. Finally, the side chains could be seen and the amino acid sequence was established. To our surprise, it is actually formed from the “stem” helices α 3, and α 4-5, instead of dagger domain helices, as had been thought

(Zhang et al. 2016). The overall model building process is depicted in Figure 3.25. There is a large part of unfurling domain sequence becoming invisible. The resolved residues with secondary structure are indicated in Figure 3.26.

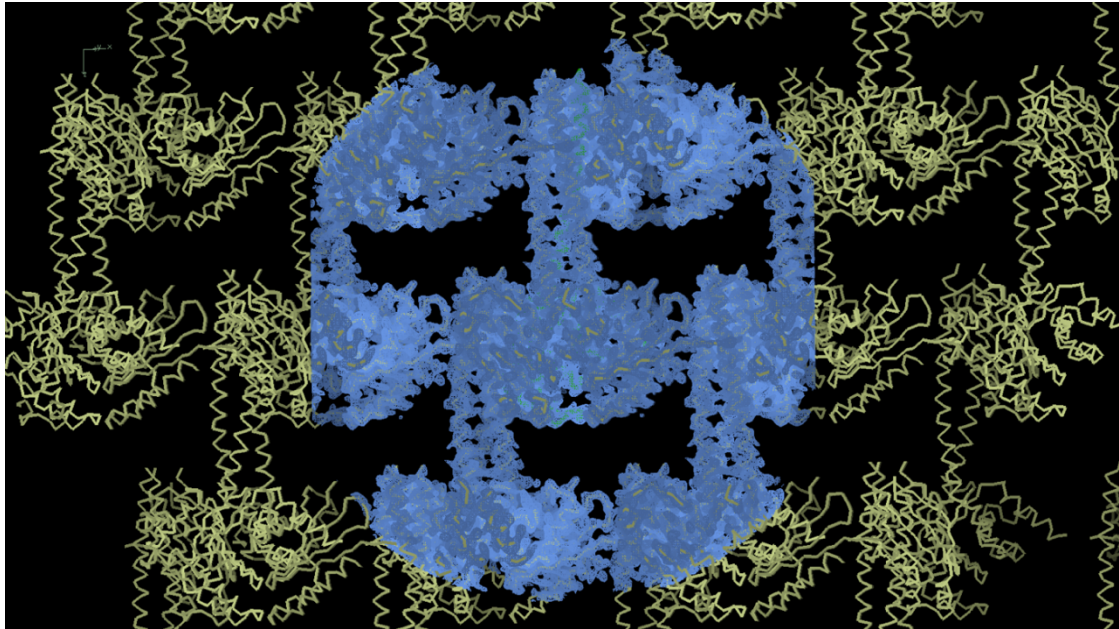


Figure 3.23 The BTV-15 VP5 low pH crystal packing and electron density map. When a poly alanine was placed in the helical density other than anchoring domain, the overall density improves. The density map is shown as blue mesh. The protein sequence Ca is shown as yellow lines.

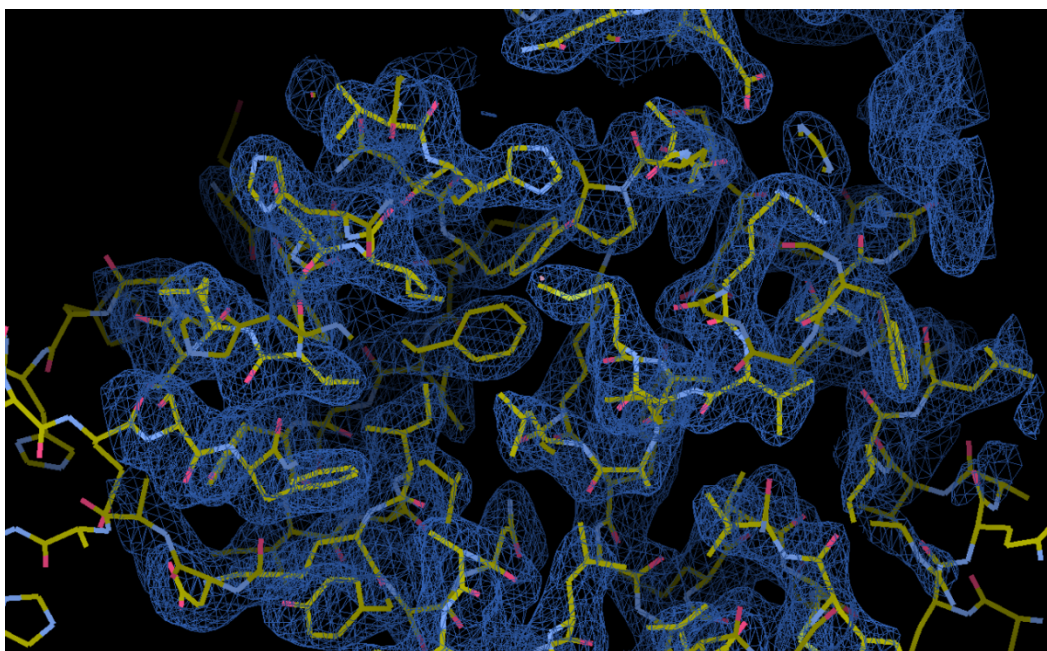


Figure 3.24 Electron density map (2Fo-Fc) shows an area of low pH VP5 structure during the model building in Coot. Contour level at 1 σ . The density map is shown as a blue mesh and amino acids are shown as yellow sticks with oxygen coloured red and nitrogen coloured cyan.

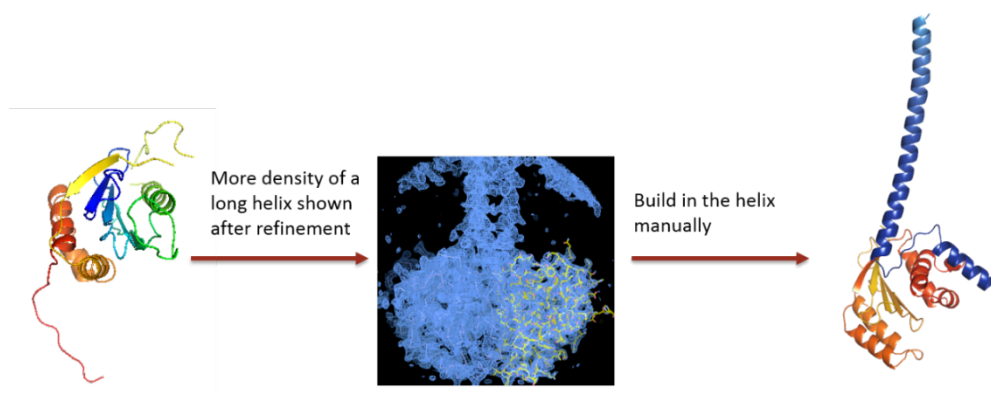


Figure 3.25 The BTv-15 VP5 low pH crystal structure refinement process as indicated by cartoon and electron density maps. The left panel shows a cartoon model of the anchoring domain (coloured rainbow, from N-terminal blue to C-terminal red), when fitted into low pH model (middle panel) there is extra density, and model was built manually into this extra helical density. The resulting model is shown in the right panel. The anchoring domain model is coloured red, the manually built unfurling domain helix is coloured blue.

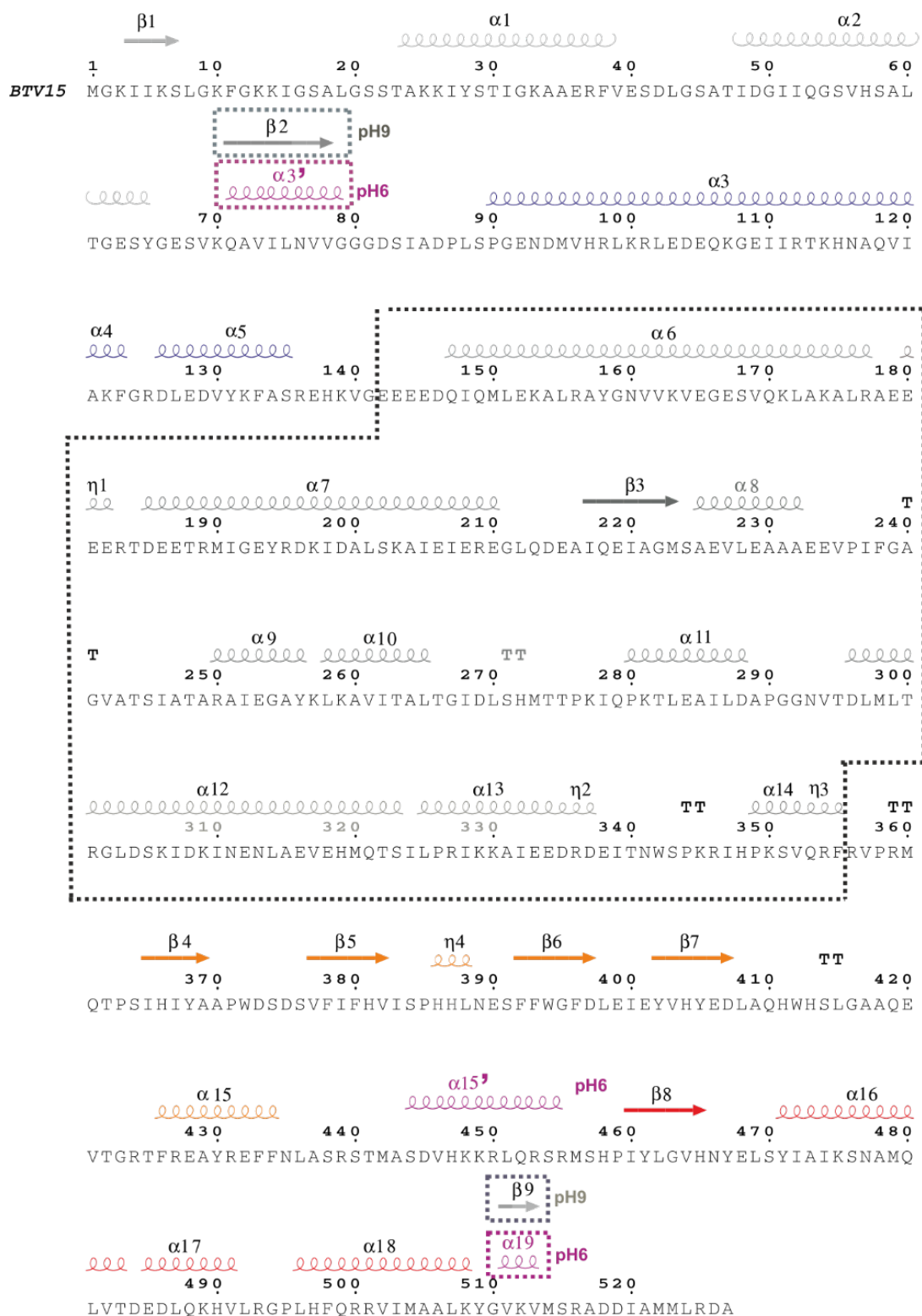


Figure 3.26 BTV-15 VP5 amino acid sequence and low pH secondary structure annotation. The secondary structure is shown above the amino acids sequence. Grey coloured elements in black dash line box indicate the corresponding part of structure is visible in the low pH structure. The purple dash line boxes indicate the secondary structures change in the low pH structure.

3.6.2. Overall structure of low pH VP5 structure

The structure was refined with repetitive cycles of Refmac5 (CCP4 package) and rebuilt using Coot before finally being refined with Phenix refine, with final R_{work} 19.9% R_{free} 21.3%, and good geometry. The refinement statistics are shown in Table 3.2.

There is only one molecule in an asymmetric unit, and it forms biological trimers with crystallographic symmetry in space group H3. The resolved amino acid sequence is shown in Figure 3.26. The cartoon presentation of the monomer and trimer is shown in Figure 3.27. The anchoring domain retains a general conformation similar to the high pH structure, especially the “meander” of β strands ($\beta 4$ - $\beta 7$) as well as $\beta 8$ strand (Figure 3.28). Helices $\alpha 15$ -18 also remain in a similar position as in the high pH, but the loop region between $\alpha 15$ and $\beta 8$ (S444-R455) becomes a nice ordered helix ($\alpha 15'$, Figure 3.28D). The most C-terminal high pH β strand ($\beta 9$) is dramatically changed to a helix ($\alpha 19$) at low pH (Figure 3.28C). The even more marked changes are in the remaining N-terminal unfurling domain. The N-terminal high pH β strand ($\beta 2$) becomes a helix ($\alpha 3'$) at low pH, and the high pH stem helix shifts position in low pH and merges the two following bent alpha helices ($\alpha 4$ - $\alpha 5$) into one long straight helix. The SHP programme used 217 residues of low pH structure to align 428 residues of high pH structure. There were 134 sequence identities out of 149 equivalences (89.9%). The Average distance is 1.74 Å and the RMS distance is 2.02 Å. The distance of corresponding C α s from both pH structures is plotted in Figure 3.29. A lot of residues show large distances up to 4.7 Å.

As the “meander” motif is relatively unchanged, the alignment of both low and high pH structures based on “meander” motif was performed, and the result shown in Figure 3.28. The low pH structure is coloured rainbow, while the high pH structure is coloured

in grey. For the clarity, the high pH parts which are invisible in low pH form are not shown.

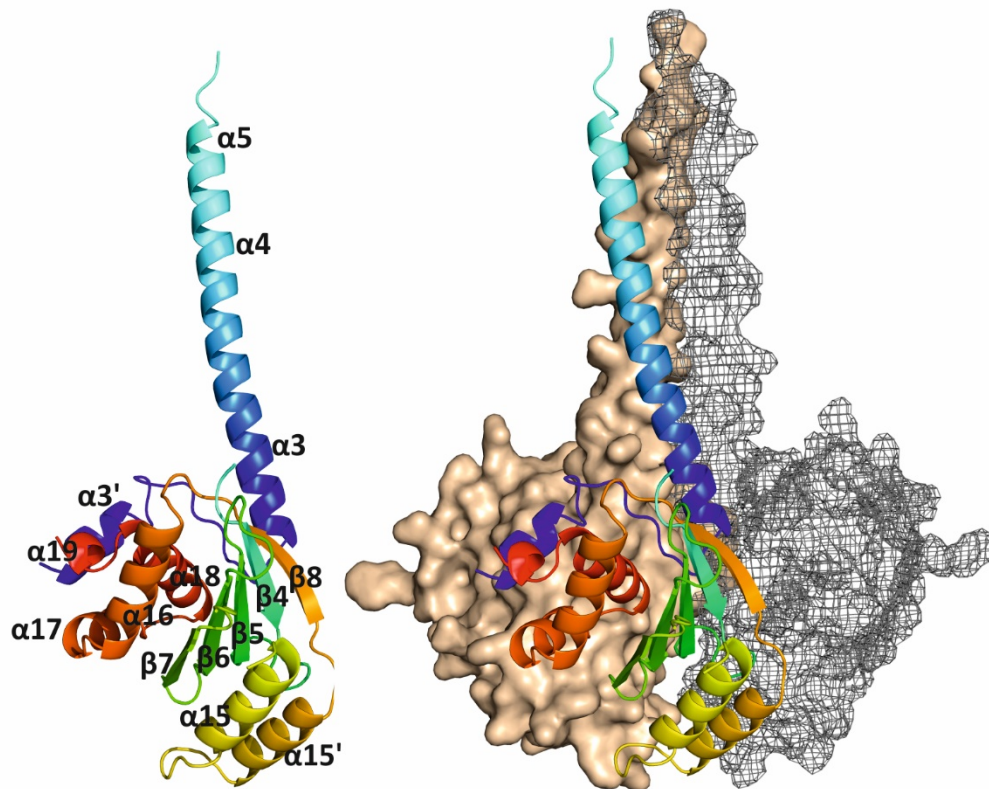


Figure 3.27 BTV15 VP5 low pH structures presented as monomer and trimers. The left panel is the cartoon representation of monomer, coloured rainbow from N-terminal blue to C-terminal red. The right panel indicate the trimeric organization of VP5. One monomer depicted as cartoon, with same colours as in monomer shown in left panel. Another monomer is shown as wheat coloured surface while the other monomer is shown as grey mesh.

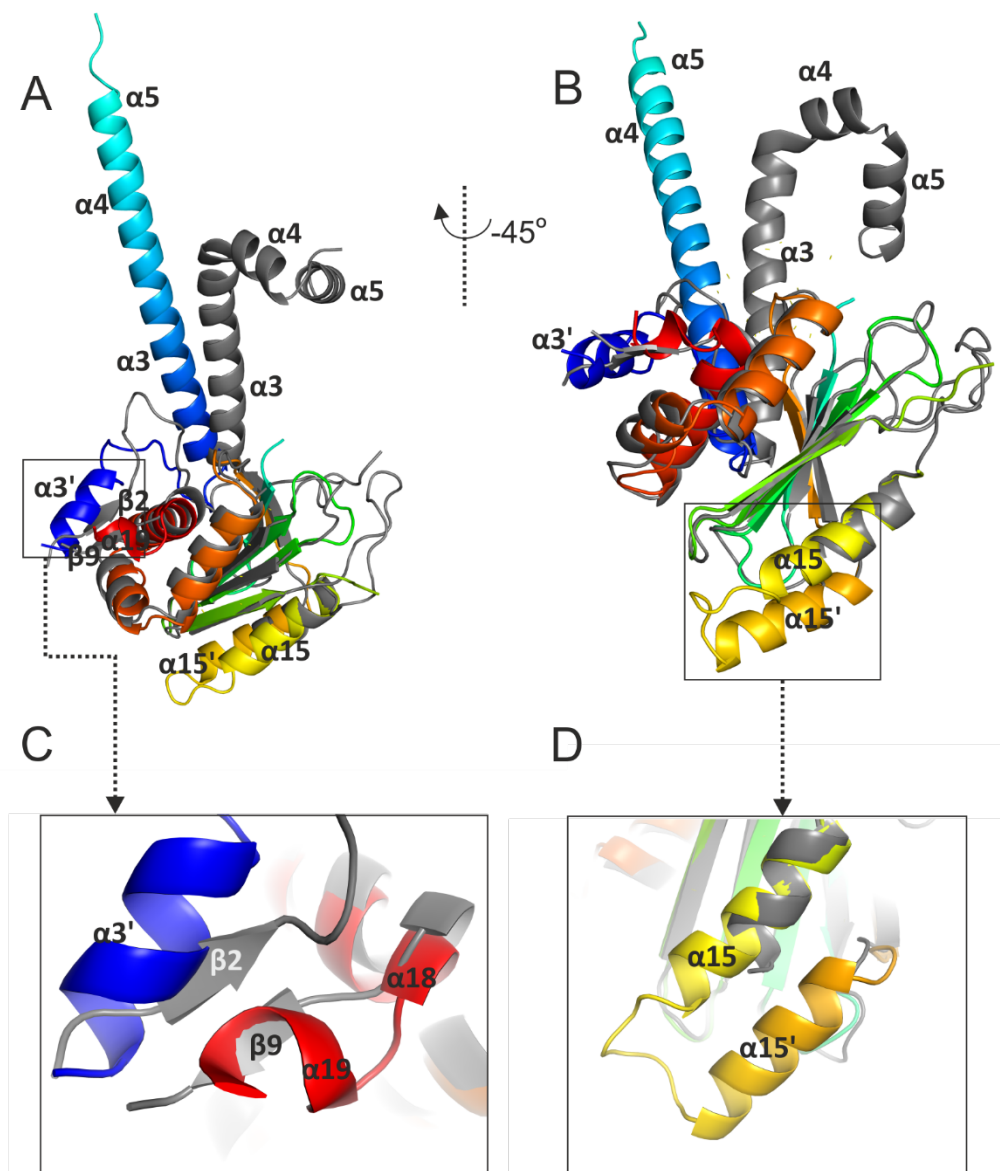


Figure 3.28 Alignment of BTV15 low pH VP5 structures with corresponding high pH structure at two different angle. The low pH structure is coloured rainbow, while high pH corresponding structures are coloured grey. The high pH unfurling domain parts which are invisible in corresponding low pH structure is not plotted. (A) is -45° rotation of (B). (C) and (D) are the close up views of the boxed area of (A) and (B) respectively.

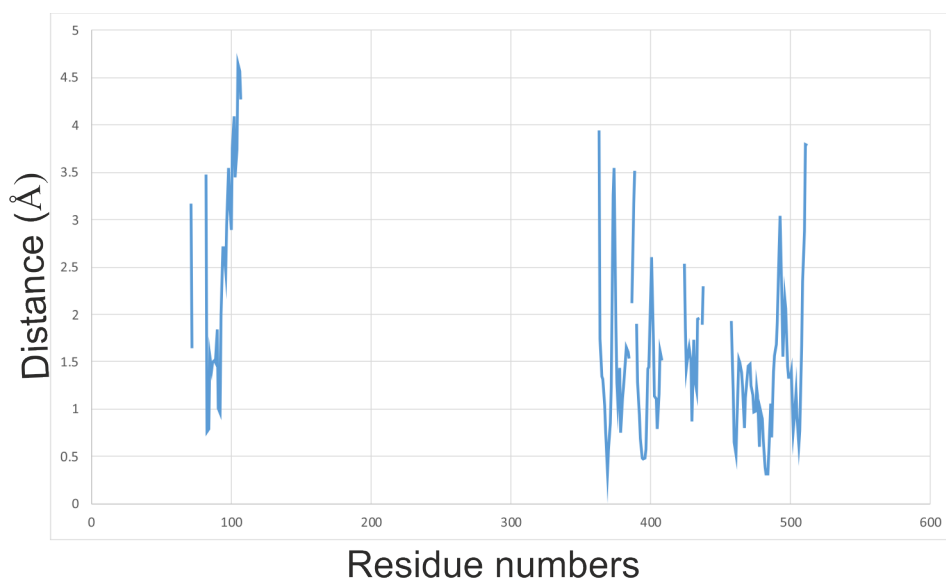


Figure 3.29 Plot of distance between corresponding C α s of the low pH and high pH structures.

3.7. Locating the pH sensor for conformation change trigger

Our low pH crystal structure revealed dramatic conformation changes as compared to the high pH crystal structure. The pH sensor for the conformational change may be composed of histidine residues. The BTV-15 VP5 monomer has 19 histidine residues, and the majority (12) are located in the anchoring domain. Figure 3.30 shows the collection of histidine residues clustered in the interface between the unfurling domain and anchoring domain. Five residues, (H365, H386, H387, H412, H465) are clustered in anchoring domain (coloured as purple sticks in the Figure 3.30). Two residues (H217 and H319) are located in the beam helix α 12 of the unfurling domain (coloured as grey sticks in the Figure 3.30), together with another two histidine residues, (H386 and H387) specifically located in the interface of the unfurling domain and anchoring domain. In addition, the conserved E316 is also in the interface. A recent experiment

suggests BTV-1 VP5 histidine cluster H384-6 is the VP5 pH sensor (Zhang et al. 2016), although without high resolution low pH data, the specific role of these residues could not be pinpointed. In BTV-15, the equivalent BTV-1 VP5 H384 position is mutated to P385. The BTV-15 VP5 H386 and H387 is equivalent to H385, H386. As both pH crystal structures at high resolution is obtained, it now becomes possible to pinpoint specific residue's conformation change.

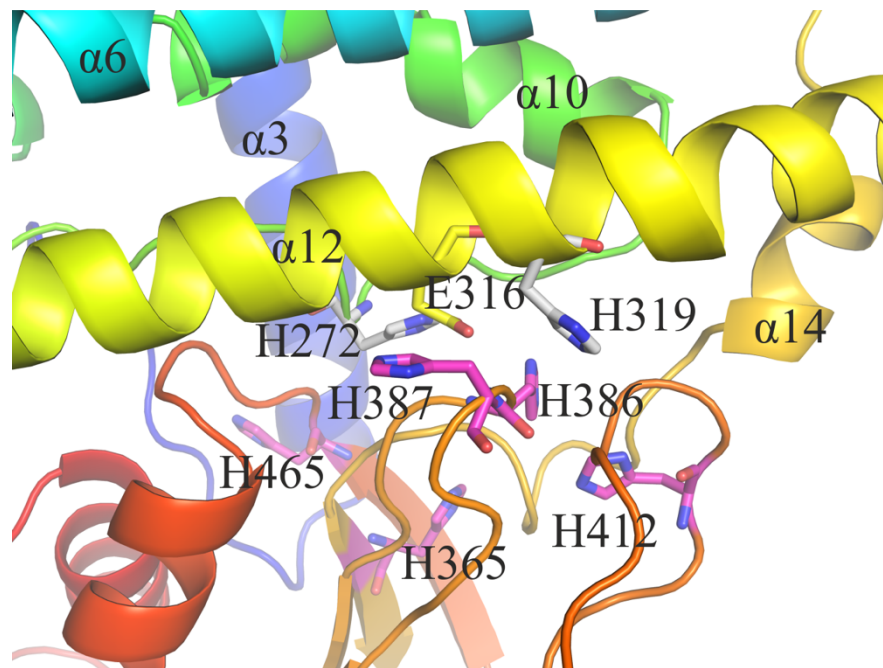


Figure 3.30 The close up view of the histidine cluster in VP5 in the area of interface between unfurling domain and anchoring domain. Histidine residues in the unfurling domain are shown as grey sticks, and in anchoring domain as purple sticks. The E316 close to the two domain interface's histidine cluster is shown as yellow sticks.

Our high pH VP5 crystal diffracted to 2.8 Å, and the low pH crystal diffracted to 2.2 Å. The histidine residues H386 and H387 are resolved well in both pH structures, as shown in Figure 3.31. The low pH 2.2 Å structure electron density around this histidine area is less well resolved compared to other parts of the structure, suggesting that this area is rather flexible. However, the side chain conformations of these histidine residues can be confidently assigned. This allows us to align these histidine residues to find out the conformational changes. The meander motifs are relatively well maintained for both high and low pH. The alignment of both meander motifs are shown in the Figure 3.32B. The histidine residues H386 and H387 located in the loop between $\beta 5$ and $\beta 6$. Further alignment of the $C\alpha$ residue of H386 reveals the conformation change of both H386 and H387. As H386 is in good contact with E316 of beam helix in high pH structure (Figure 3.32A), the alignment of low pH H386 clearly demonstrates that this histidine residue would have clashed with the E316 if the beam helix $\alpha 12$ were not moved. (Figure 3.32C). This prompted us to propose a model that low pH triggers the H386 conformation change, pushing E316 of the beam helix $\alpha 12$ and leading to movement of the helix and the majority of unfurling domain (Figure 3.32C and D). The possibility cannot be ruled out that other histidine residues may also contribute to this unfurling domain beam helix push force. For example, the contact interface residues H272 and H319 from the beam helix are not visible in the low pH structure and thus it cannot be known if they may also change conformations. And those residues may contribute to unfurling domain folding, as mutation of H272 was found to be detrimental to VP5 expression (Zhang et al. 2016).

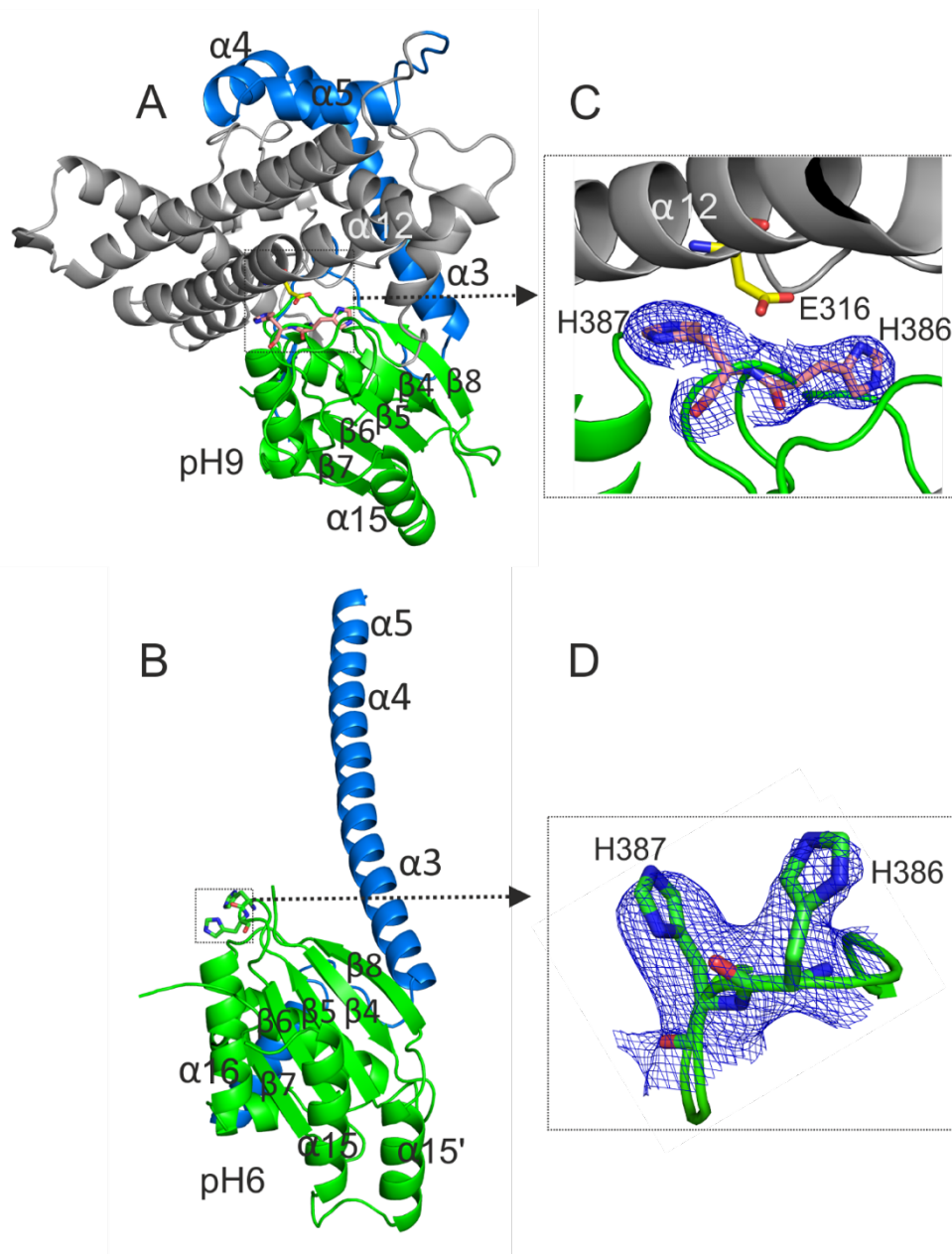


Figure 3.31 The potential pH sensor Histidine residues H386 and H387 in VP5 structures at different pHs. **(A)** is the high pH structure, shown as cartoon. The N-terminal unfurling domain is coloured blue and the rest is coloured grey. The meander domain is coloured green. **(B)** is the close up of the dashed line box area of **(A)** showing Histidine H386 and H387 residues in pink sticks with electron density 2Fo-Fc map as blue mesh contoured at 1σ . The residue E316 from $\alpha 12$ of unfurling domain is shown as yellow sticks, indicating if H386 changes its side chain conformation then it is likely to clash with E316. **(C)** displays the low pH structure shown as cartoon, coloured same as in **(A)**. **(D)** is the enlarged dash line box area of **(C)** showing Histidine H386 and H387 residues in green sticks with electron density 2Fo-Fc map as blue mesh contoured at 1σ .

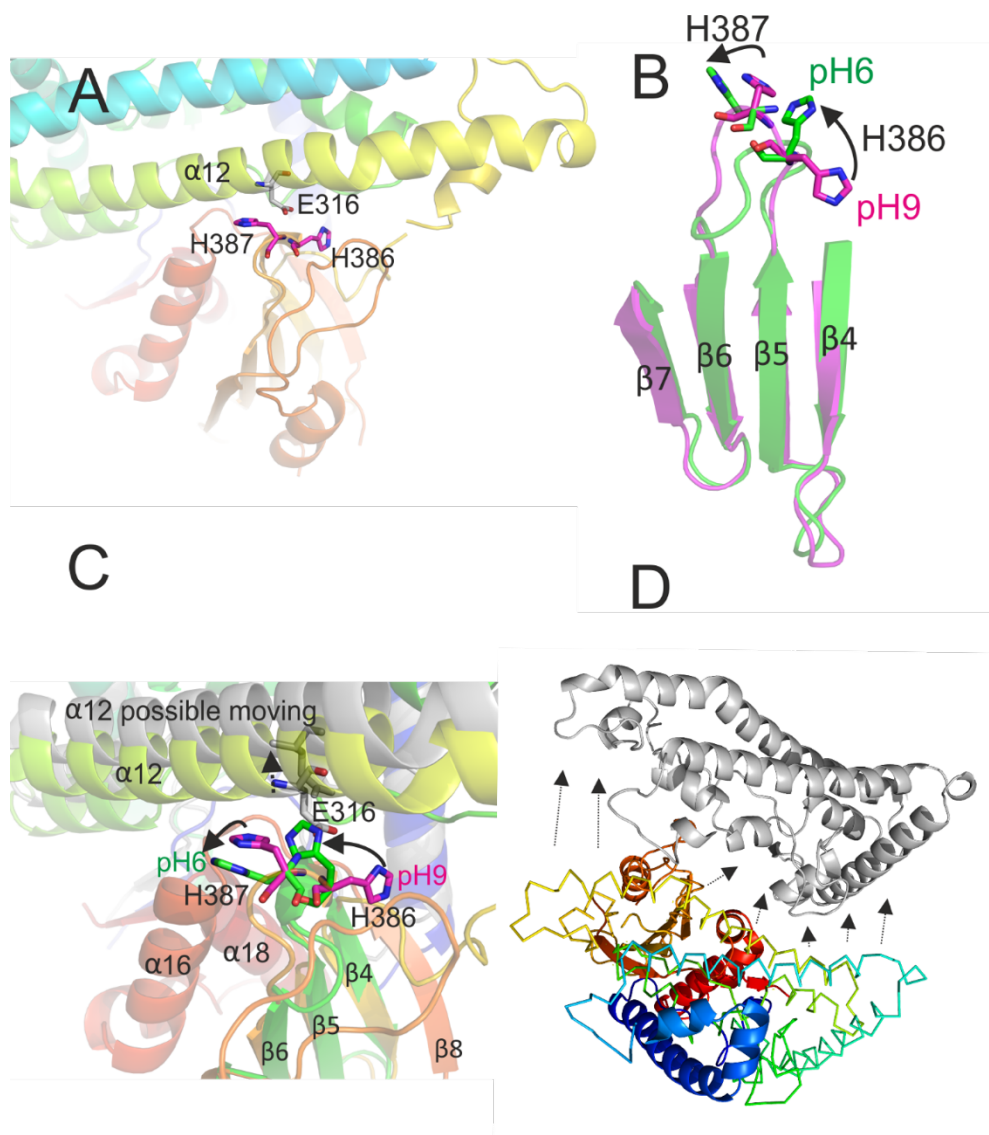


Figure 3.32 The proposed mechanism of low pH triggered VP5 unfurling domain disorder or detachment. (A) The histidine residues H386 and H387 in high pH structure are shown as purple sticks, with the unfurling domain $\alpha 12$ residue E316 shown as grey sticks. (B) Alignment of VP5 meander domains from high pH (coloured purple) and low pH (coloured green). Histidine residues H386 and H387 sidechain conformation changes are indicated. (C) Similar view of (A) when the low pH meander domain aligned to high pH structure. It is indicating that when H386 is changed to low pH conformation, it would clash with the high pH unfurling domain residue E316. This may lead to movement of $\alpha 12$ (coloured as grey). (D) Proposed diagram shows the potential movement of unfurling domain parts which is disordered or detached in low pH structure. The low pH parts that remain are shown in coloured cartoon, with the low pH invisible parts as coloured ribbon. Black arrows indicate the potential direction of movement, and the grey cartoon showing moved parts which are invisible in the low pH structure.

3.8. Fitting the low pH crystal structures into cryo-EM map.

X-ray crystallography was used by us to solve the low pH VP5 structure at high resolution. The structure revealed distinctive features, such as dis-ordered/detached parts of the unfurling domain and a protruded stem helix. As our crystal structure is from the insect cell expressed single protein, it would be interesting to know if the features observed here are physiologically relevant in the context of the complete virus. The high resolution low pH crystal structure was tried to fit into the recently available 9 Å low pH cryo-EM density (Zhang et al. 2016). Indeed, our low pH crystal structure fits well into the cryo-EM map, as shown in Figure 3.33. The fitting average score is 0.815 ± 0.02 . This confirms that our high resolution crystal structure is biologically relevant, and that the distinguishing features we observe may be genuine.

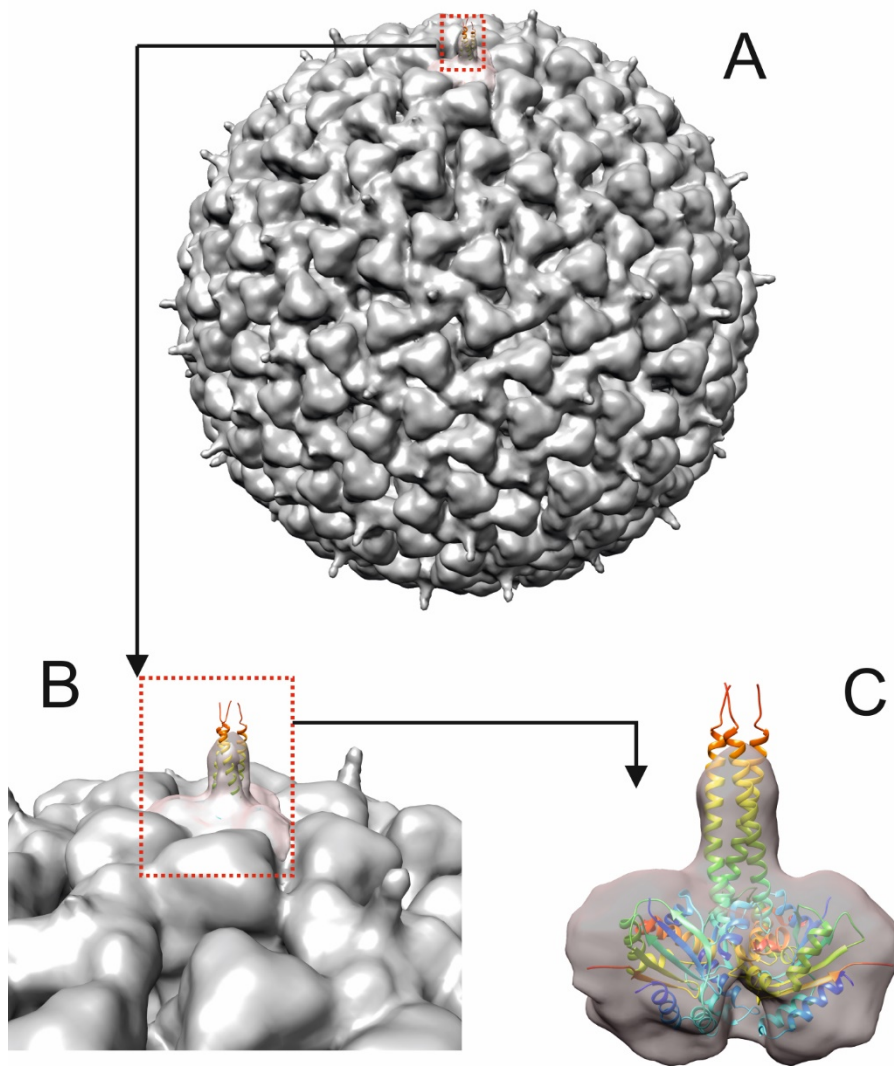


Figure 3.33 Low pH BTV15 VP5 crystal structure fitted into low pH cryo-EM structures. (A) shows the BTV whole virus structure at low pH. (B) is the close up area of (A) for VP5 trimer. (C) is a close up view of (B) showing VP5 trimer density and fitted crystal structure. The low pH crystal trimer structure is shown as coloured cartoon and cryo-EM map is coloured grey. The fit top score is 0.86 and average score is 0.815 ± 0.02 .

3.9. Discussion

X-ray crystallography was used here to solve the BTV-15 VP5 structures at both high and low pH. Our high resolution crystal structures (2.8 Å and 2.2 Å for high and low pH respectively) give more precise information than is provided by the cryo-EM structures. The cryo-EM structures are at 3.5 Å and 9 Å resolution for high and low pH respectively. Our low pH structure reveals a lot of new features of which 9 Å low resolution cryo-EM could not resolve. For example, the $\beta 2$ and $\beta 9$ strands dramatically change their conformation to alpha helices ($\alpha 3'$ and $\alpha 19$). A comparison of high and low pH structures also revealed the pH sensor histidines (H386 and H397) changes conformation. Most importantly, the low pH structure revealed the conformation change of stem helix $\alpha 3$ which fuses with $\alpha 4-5$ to form a long straight helix sticking out of the molecule. This represents the first time such a conformation has been seen in a non-enveloped virus, and suggests a mechanism of “penetrating” for this protein. In the cryo-EM structure, due to its low resolution, although an object could be seen sticking out it was wrongly suggested to be a dagger domain helix (Zhang et al. 2016).

In our low pH structure, one mystery remains. The protein used for crystallization contains both unfurling domain and anchoring domain. We didn't treat the protein with any protease, however, the majority of the unfurling domain is invisible from the structure. It is possible that the majority of the unfurling domain is still packed in the crystal but just completely disordered. The relatively small unit cell is unlikely to be able to accommodate all the unfurling domain and anchoring domain sequence if they were well ordered. There are two hypotheses could be explored. One is the “invisible” sequence still packed within the crystal, but just completely disordered, and the disordered part “resemble” solvent content. From the crystal packing, as shown in

Figure 3.26, it is clear that there are some “unstuffed” areas, and it is conceivable that the disordered part could be packed in this area. Theoretically, this could be established by mass-spectrometry of protein derived from the low pH crystals. However, only a single crystal was produced, which could not be repeated.

Another possibility is that the “invisible” part was not present in the crystal. Some residual protease cleaved the “invisible” part of unfurling domain. The BTV-15 VP5 has a predicted furin site at residue 98-102 (RLKRL). The furin (or PACE, paired basic amino acid cleaving enzyme) cleavage site (R-X-[RK]-R-|-X) is frequently used for enveloped virus fusion process, for instance HIV-1 uses furin to process GP160 precursor into GP120 and GP41 (Hallenberger et al. 1990), and in Ebola virus, the glycoprotein is processed into GP1 and GP2 by furin (Mulherkar et al. 2012). However, the BTV-15 VP5's putative furin site is located at the “stem” helix α_3 , which is ordered in both crystal structures at low or high pH. Thus, it suggests this putative furin cleavage site may not be used. The VP5 protein used for crystallization was produced in insect cells, and the protein is of predicted size about ~55 kD as in

Figure 3.5, thus further suggesting it is not processed by furin.

BTV-15 VP5 also has a predicted caspase-3 and caspase-7 cleavage site at residue 196-200. The sequence DKIDA fits the cleavage pattern [[^]P][[^]DEWHFYC]D[GSAN]. This putative cutting site is located in unfurling domain α_7 . This part of sequence is visible in high pH crystal structure, but it is not visible in low pH structure. Some enveloped viruses, like Ebola use caspase L/B to process glycoprotein 1(GP1) in late endosomes to expose the receptor (Niemann-Pick disease, type C1, NPC1) binding site (Mulherkar et al. 2012). In the non-enveloped viruses, like reovirus, the coat protein σ_3 can be processed by many kinds of endo/lysosomal proteases. The cut sites are located in solvent exposed loops or arches (Jané-valbuena et al. 2002; Mendez et

al. 2003). In the grass carp reovirus, VP5 is also processed by proteases. Virus particles treated with trypsin or chymotrypsin can have an enhanced infectivity. Although we may not know for certain if in the invisible part of unfurling domain is proteolytic detached, but it provided an important clue that similar proteolytic process may be also possible in BTV VP5. This warrants further investigation using advanced mass-spectrometry techniques in the real physiological context.

If future experiments prove that proteolytic processing also happens in BTV VP5, we can propose a structural model as shown in Figure 3.34. The unfurling domain parts other than stem helices are proteolytically processed but remain loosely attached to the anchoring domain. When the pH drops further, the pH sensing histidine residues get protonated and H386/H387 changes conformation. This moved histidine clashes with E316 in the main contacting unfurling domain helix $\alpha 12$, which causes the majority of the unfurling domain to fall off the anchoring domain, and allows the conformation of stem helix $\alpha 3$ and the following helices $\alpha 4$ -5 to change conformation, and become one long straight helix sticking out the molecule.

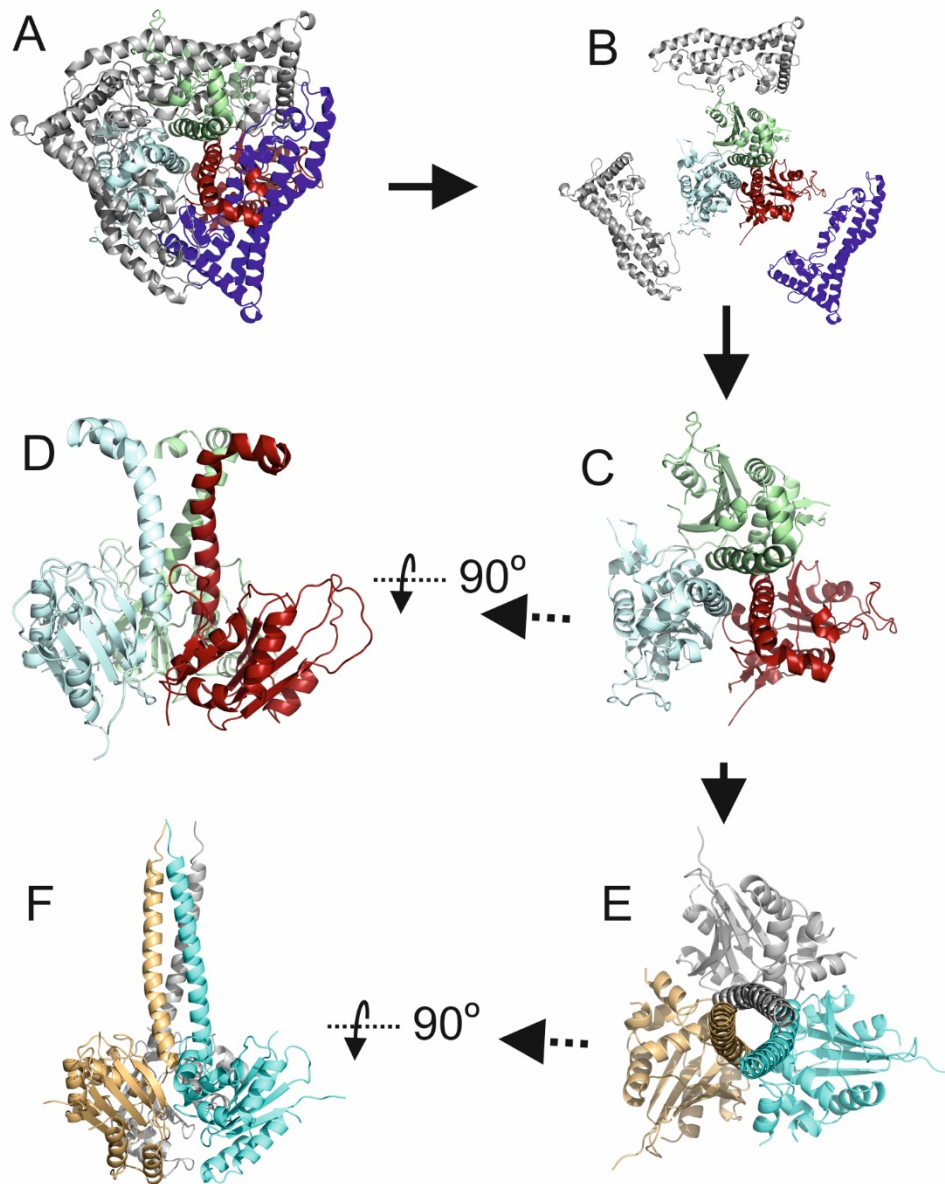


Figure 3.34 Proposed cartoon model of conformational changes of BTV VP5. When the VP5 environmental pH drops, unidentified proteases may proteolysis some residues in the unfurling domain. (A) represents a trimer, with main unfurling domain coloured blue in one copy and grey for the other two. The anchoring domain and stem helix $\alpha 3$ and following $\alpha 4-5$ were coloured dark red, light green and beige for each monomer respectively. (B) indicates the main unfurling domains falling off. (C) and (D) are different views of the anchoring domain and stem helices before conformational change and (E-F) have the same view of (C-D) after conformational change. Each monomer is coloured grey, cyan and light yellow.

Animal viruses are categorized as enveloped and non-enveloped. Enveloped viruses use glycoproteins imbedded in a cell derived lipid bilayer to fuse with the host cell membrane and deliver viral genetic material into the host cells. For example, human immunodeficiency virus type 1 (HIV-1) uses an envelope glycoprotein precursor gp160 which is processed by cellular furin endoprotease to generate stable gp120 and gp41. HIV trimeric gp120 is associated non-covalently with the gp41 transmembrane protein. Receptor binding to gp120 triggers conformational changes in both gp120 and gp41. The gp41 forms a trimer of α -helical hairpins and triggers the membrane fusion.

The membrane penetration process of non-enveloped viruses may be fundamentally different from enveloped viruses, but it appears to follow a general principle that the virus traffics to a penetration site and engages cellular factors that impart structural changes to enable the viral particle to bind and penetrate a target membrane. The molecular details of membrane disruption and the structural characteristics of the domains involved in non-enveloped viruses are not as well understood. Our low pH structure revealed a unique feature of sticking out long stem helix. The three protruding helices from a trimer, are most likely responsible for penetration of endosomal membrane.

When the protruding stem helices from a trimer structure was submitted to Dali server (http://ekhidna.biocenter.helsinki.fi/dali_server) to search for similar structures, the top hits are HIV GP41 helices. The top hit alignment is shown below.

DSSP	-LHHH-----	
Query	-GENDMVHRLKRLDEDEQKGEEIIRTKHNAQVI AKFGRDLEDVY-----	41
ident		
Sbjct	qLLSGIVQQNNLLRAIEAQQHLLQLTVWGIKQLQARILAGgttwmewdreinnytslih	60
DSSP	LHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHL Lllhhhhh hhhhhhhhh	
DSSP	-----hhlLHHH	
Query	-----kfaPGENDMVHRLKRLDEDEQKGEEIIRTKHNAQVI AKFGRD	81
ident		
Sbjct	sleesqnqqekneqelle gsggQLLSGIVQQNNLLRAIEAQQHLLQLTVWGIKQLQAR	120
DSSP	hhhhhhhhhhhhhhhhhh llllLHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	
DSSP	H-----hbbbhLHHH	
Query	LE-----dvykfapGENDMVHRLKRLDEDEQ	106
ident		
Sbjct	iLghttwmewdreinnytslih sleesqnqqekneqelle gsggQLLSGIVQQNNLLRAI	180
DSSP	HLlhbbbbbbb bbbbbb bbbbbb bbbbbb bbbbbb bbbbbb llllLHHHHHHHHHHHHHHHHH	
DSSP	HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH	
Query	KGEIIRTKHNAQVI AKFGRDLEDv y k f a	134
ident		
Sbjct	EAAQHLLQLTVWGIKQL QARILA----	203
DSSP	HHHHHHHHHHHHHHHHHHHHHHHHHHHHH LL-----	

a b c d e f g
H P P H C P C

148

glycoproteins use low pH triggered heptad repeat (HR1) forming a helix bundle as part of the fusion mechanism.

It is intriguing that our non-enveloped BTV VP5 low pH protruding helix has some resemblance of enveloped HIV-1 gp41 heptad repeat. The alignment of the low pH BTV VP5 extended stem helices with the gp41 HR1 is shown in Figure 3.35. This phenomenon may suggest some resemblance between the mechanisms of overall viral protein membrane penetration between enveloped and non-enveloped viruses.

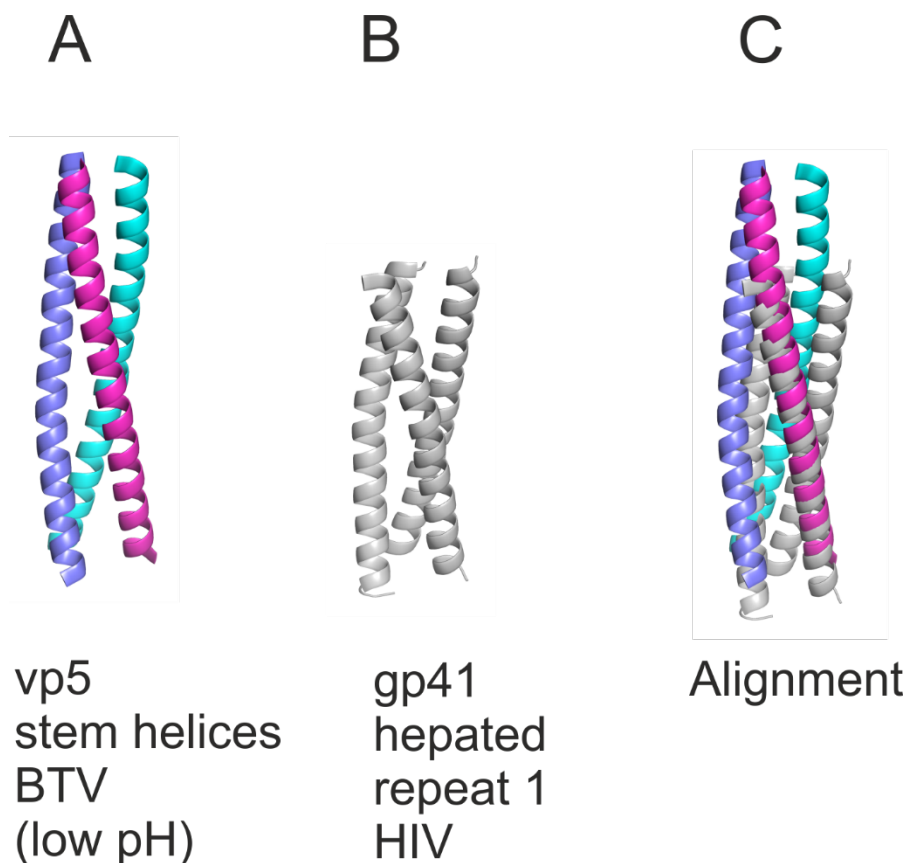


Figure 3.35 A cartoon presentation of BTV15-VP5 stem helices at low pH (A) and HIV-1 gp41 heptad repeat 1 (B) and their alignment (C).

Chapter 4.

Structural Studies of MSV

4.1. Structural Studies of MSV VLPs

4.1.1. Construct design and cloning

The vector pOPIN series created by the Oxford Protein Production Facility (OPPF) was used for cloning of MSV CP gene with ligation-independent cloning (LIC). The pOPINE vectors used here to facilitate his tagging at C-terminal of the protein. A detailed vector map of the pOPINE is shown in Figure 3.3. The construct for cloning and expressing is the full length MSV CP with KHHHHHHH for C-terminal tag (from pOPINE vector). The size of the whole construct is 28 kDa. Genes were then cloned into the appropriate pOPIN vectors using the In-Fusion system as described in Chapter 2.

4.1.2. Protein expression and purification of MSV-CP

The purification procedure including cell breaking, centrifugation, affinity column purification, size exclusion gel filtration with S200 column. followed by a final ultracentrifugation with a 10-30% sucrose density gradient were described in the methods. The spectrometry result and the SDS-PAGE analysis of the gel filtration step is shown below (Figure 4.1). As seen from the result, the peak for the MSV-CP is a single peak, but with some contamination in the latter half of the peak. To best guarantee the purity of the sample, only the first four fraction were pooled together to be carried on to further procedures. One feature of the gel filtration result worth mentioning is that the absorption at the wavelength of 260 nm, represented by the A260 Y-axis ordinates, is significantly higher than a normal protein purification ($A_{260} \approx 1/2 A_{280}$), the absorption of the A260 is twice as much as that of A280. This may be caused by the encapsidation of nucleic acid inside the VLPs, resulting in a high A260/A280 value. The virus was first checked for optimal concentration and quality

control in negative stained EM. The sample preparation after sucrose gradient was stained with Uranyl Formate and examined for quality with a Tecnai T12 microscope. For the detailed procedure see the Materials and Methods chapter. Negatively stained MSV-VLP particles exhibit characteristic icosahedral morphology (Figure 4.2). The virions showed a rather heterogeneous distribution of particles with particle size of $200 \pm 10 \text{ \AA}$, and some pentamers can be seen in the background. Overall the particles were not aggregated, and even though some particles had dissociated into pentamers, the majority of the particles were intact virus even in a diluted sample.

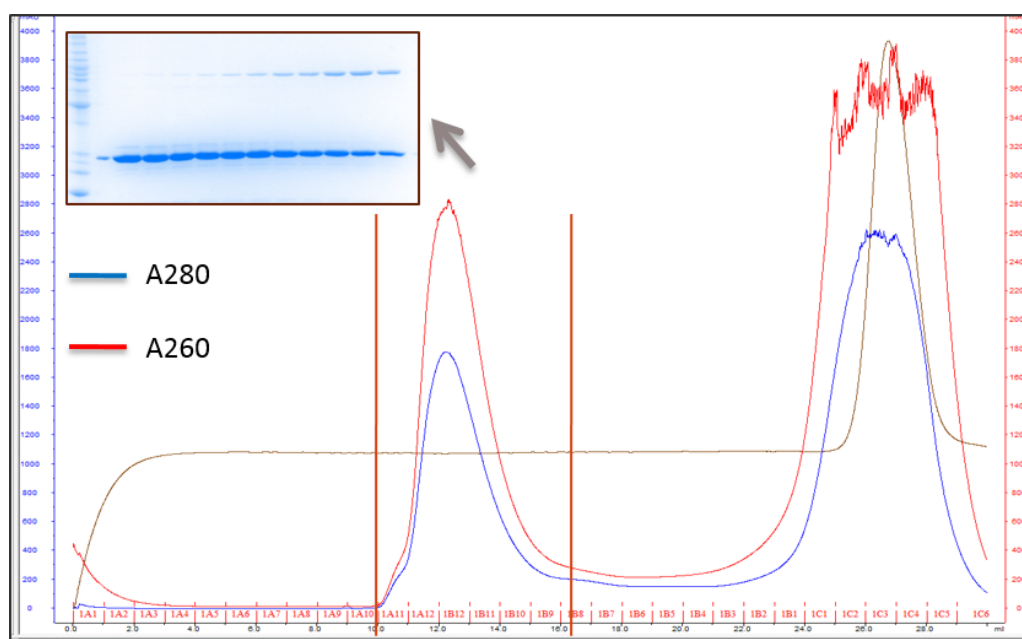


Figure 4.1 The spectrometry result and the SDS-PAGE test for the gel filtration of MSV CP. The left Y-axis ordinates records for absorption of A280, representing protein concentration. The right Y-axis ordinates is for A260, representing nucleic acid concentration.

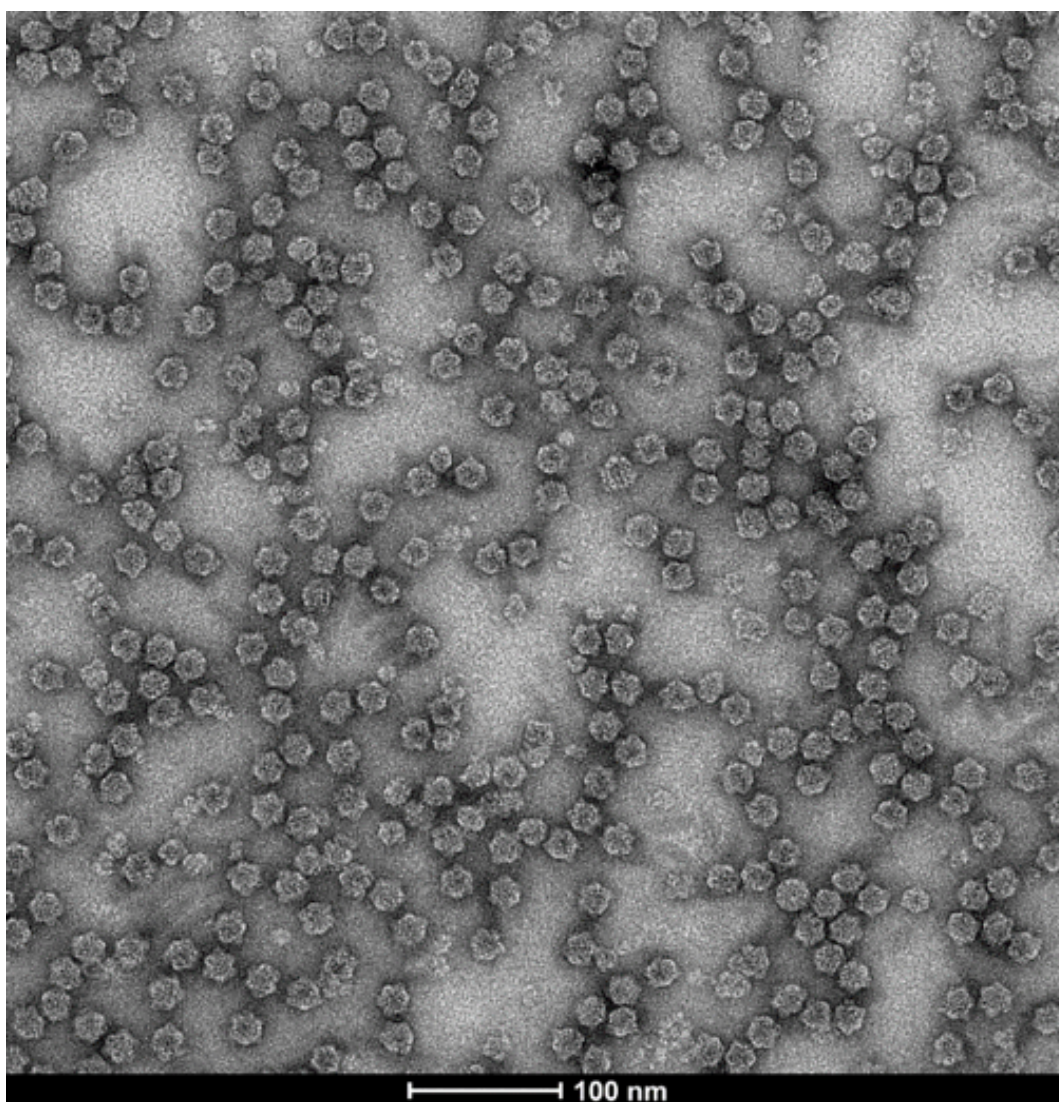


Figure 4.2 MSV-VLP checked with negative stained EM. The sample was taken after sucrose gradient and stained with Uranyl Formate. Negatively stained MSV-VLP particles exhibit characteristic icosahedral morphology.

4.1.3. Cryo-EM structure of MSV-VLP

Once the quality of the preparation was verified by negative-stained EM, the sample was diluted to 0.5 mg/mL, to give a good particle density on the plunge-frozen grids. This value was about 10-20 times that of the optimal concentration to be examined with negative-stain EM. After plunge-freezing, the virus particles were imaged with

cryo-EM, on a Polara microscope (FEI). The detailed imaging procedure was described in the Materials and Method chapter. After Motioncorr and CTF correction, particles were picked manually. Roughly 5000 particles were subsequently classified into 40 2D classes (Figure 4.3). Relion was used for classifying the picked particles iteratively based on 2D projection. Classes outlined with a red box in Figure 4.3 were chosen to carry on with further processing. These 22 classes contained some 4600 out of 5000 particles.

Icosahedral Symmetry were applied to the reconstruction. The particles were then classified in 3D into 20 classes with Relion, using the STNV-CP fitted MSV crystal map (see Materials and Methods for details) as a template that filtered to a very low resolution of 50 Å so as not to bias the final structure. The 3D classification result is shown as Figure 4.4. Once the 3D classification had converged, Relion also provided the angular accuracy of rotation for the determination of each 3D class-average. However, the determination of which 3D classes were to be further refined to give the final structure was done by visual inspection of which class or classes looks the best. Figure 4.4 shows the 2D central slice through each of the 3D classes obtained for the MSV-VLP particles, Class 1 on its own contained 2067 particles, and is the best among all classes judged from Chimera visualisation. This class of reconstruction displayed higher-resolution features more clearly and was selected for further structural refinement. This 3D reconstruction result from Class 1 was again used as a model to do another round of 3D classification against all particles selected from the 2D classes, and for this time among the 20 2D classes, the 14th Class (boxed in red in Figure 4.3) was the best which consisted of 1453 particles. The structure was further refined with Relion refinement and post_process (so-called particle polishing), to produce the final structure. 1453 particles contributed to the final structure of MSV-VLP and the

final resolution was determined to 9.5 Å. This particle structure, along with the particle sliced by half, and a central slice of it are shown in Figure 4.5 to show the density. The MSV single particle has a diameter of 200 Å across the fivefold axes of symmetry, suggesting a tighter configuration of the pentameric capsomers than in the geminate particle. The single capsid contains only one type of pentameric unit, and it obeys strict five-fold symmetry as imposed by the icosahedral symmetry of the particle.

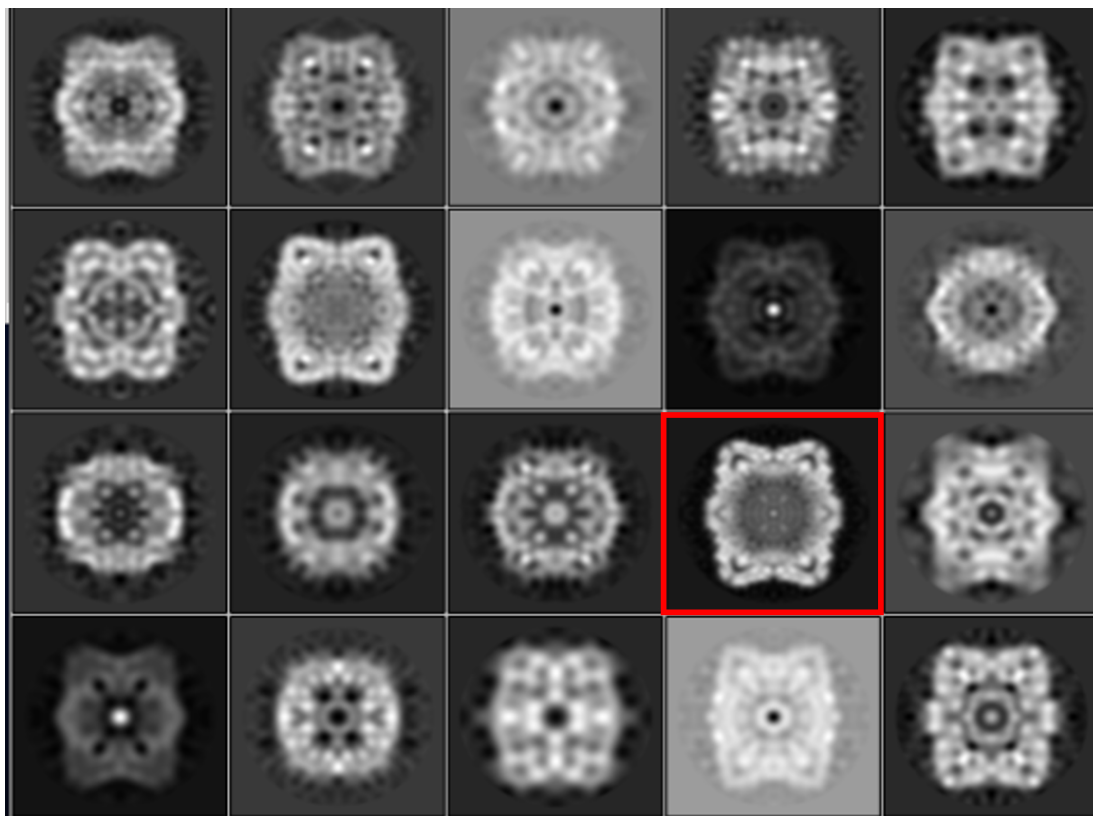


Figure 4.3 Central slices through the 2D classes of the MSV-VLP particles. .Red boxes indicate classes taken forward to 3D classification.

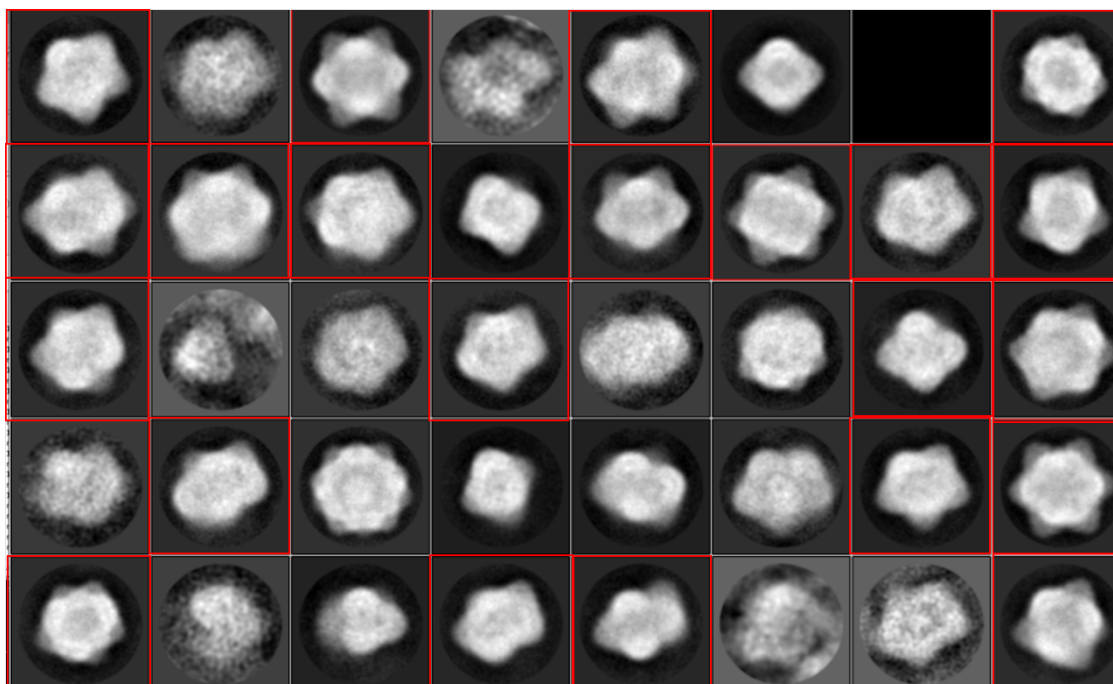


Figure 4.4 2D projections of the 20 3D classes of the MSV-VLP particles. Red boxes indicate the class (Class 14) proceeded to refinement.

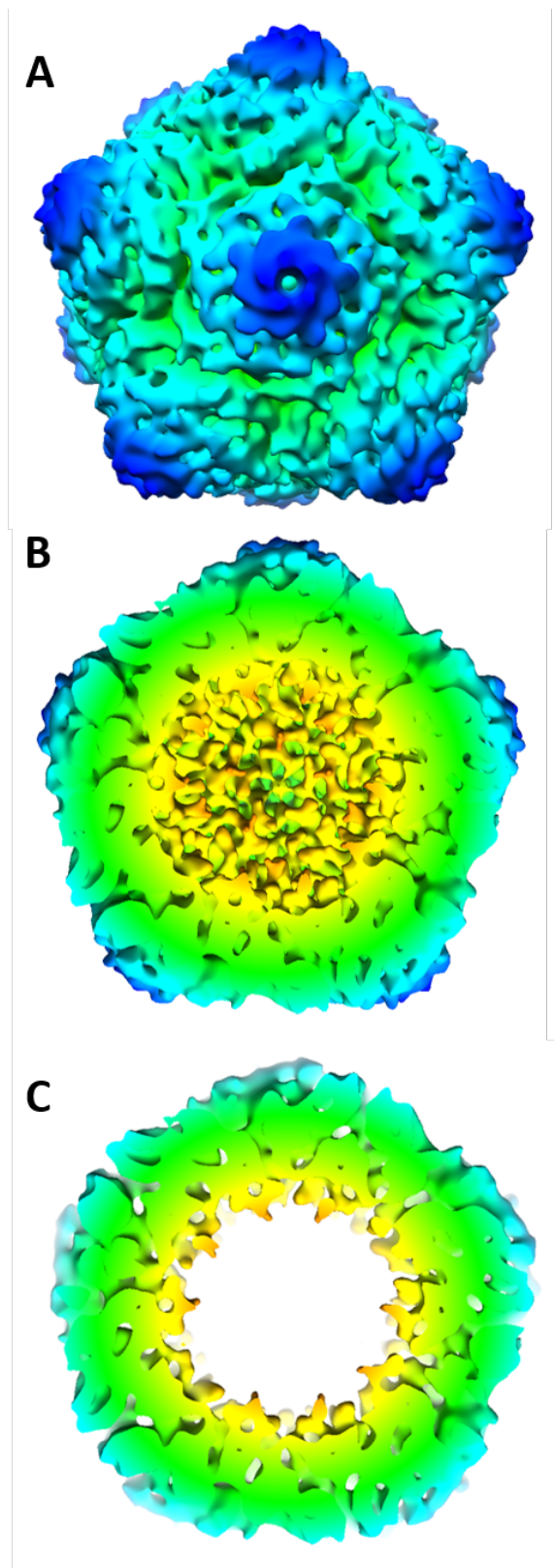


Figure 4.5 Final structure of MSV-VLP with a final resolution determined to 9.5 Å. Full particle (A), half particle clipped from the centre (B) and a central slice (C) are shown.

4.1.4. Alignment of MSV and STNV and Fitting of STNV CP into MSV VLP

As there is no known structure for MSV or other geminiviruses, we searched for the most similar structure in the protein data bank to fit into the density of the EM density. The hidden Markov method of HHpred (Hildebrand et al. 2009) was applied and the sequence found to be most similar was that of the capsid (CP) of STNV (PDB 4BCU). The alignment of MSV CP to STNV CP is shown in Figure 4.6. The structure of the STNV CP was therefore used as the best model to be fitted into the EM density of MSV. The MSV CP and STNV CP share 13% sequence identity. The atomic model of STNV CP was docked into the final cryo-EM density map of MSV VLP pentamer, using Chimera (Figure 4.7). During fitting, 10 solutions were searched for and the best fitting solution (the one with the highest cc value) was adopted. The size of the MSV CP is 27 kDa, larger than that of the STNV, which is 18 kDa, so that unaddressed density was shown in the VLP compared to the STNV CP molecule. Extra density can especially be seen close to the axis. For the best fitting solution (Figure 4.7), a correlation coefficient (CC) value of 0.73 was achieved which indicates the fitting is appropriate yet not perfect, since the size of the protein is one third bigger than the fitting model.

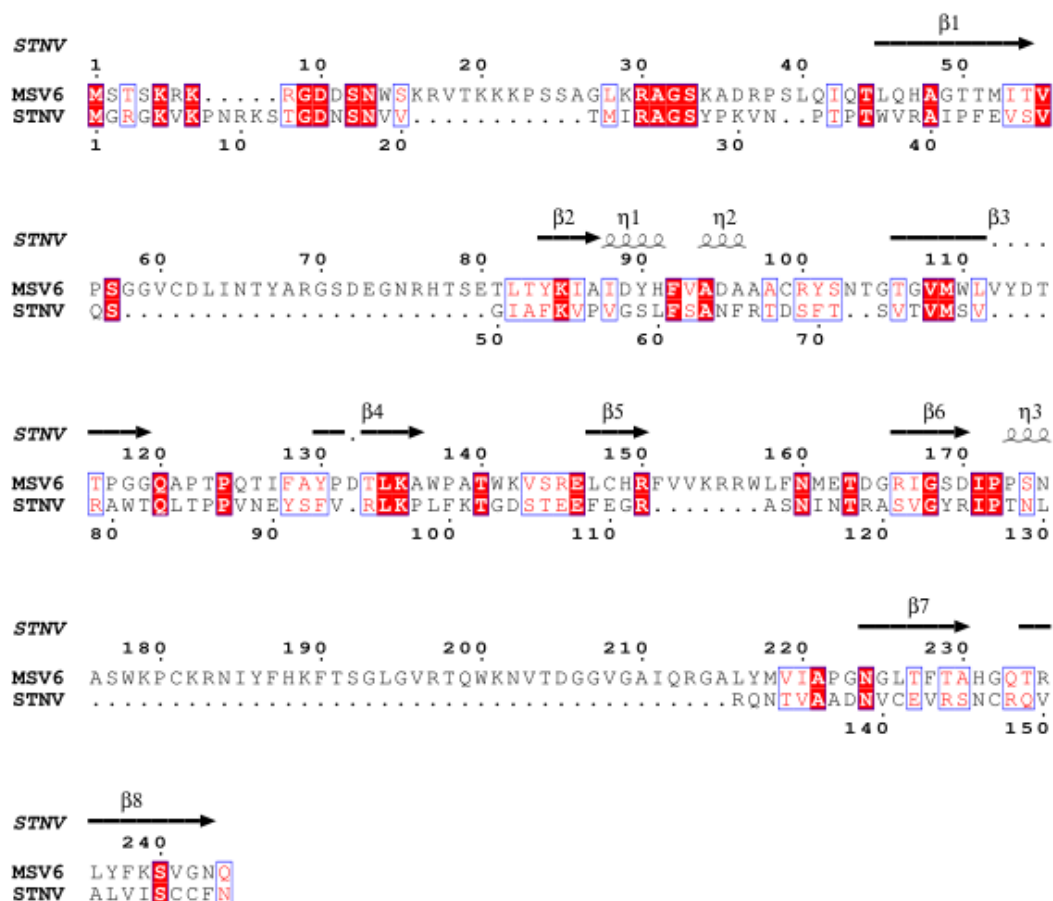


Figure 4.6 Sequence alignment of MSV CP and STNV CP. .Sequence MSV6 CP was aligned based on the sequence of STNV CP using online T-Coffee multiple sequence alignment software (<http://www.ebi.ac.uk/Tools/msa/tcoffee/>) (Ridgeway & Nw 2000) and then image prepared wit ESPrpt.

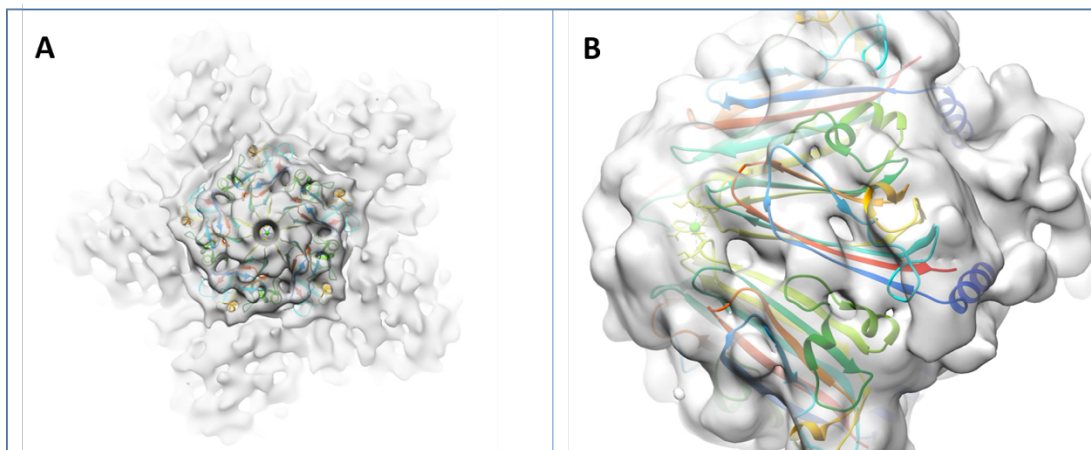


Figure 4.7 Fitting of one pentamer of the STNV CP into the EM density of a CP pentamer of MSV VLP. (A) shows a top view of the fitting from a five-fold axis and (B) is a zoom in side view of the fitting, focusing on the jelly roll structure.

4.1.5. Resmap analysis of MSV-VLP EM structure

ResMap (Kucukelbir et al. 2014) was used to visualise the local resolution of the EM 3D density maps (Figure 4.8). The areas around the connection between two pentamers show higher resolution (more of colour blue for resolution higher than 4 Å and cyan for 5 Å), while the centres of the pentamers, which are the vertexes of the icosahedral particles shows lower resolution (colour grey for resolution lower than 9 Å). This may be due to the flexibility towards the vertexes of the particles while the joints between pentamers fixed the structures in the areas more tightly.

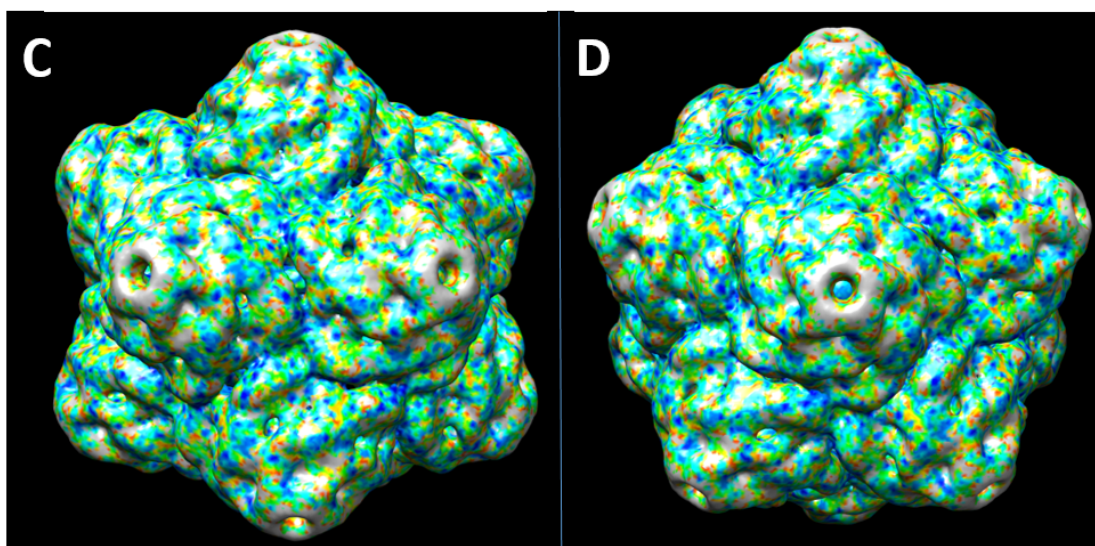


Figure 4.8 ResMap analysis of MSV_VLP to colour-code structure based on local resolution. The colours that represents for resolution ranges are: Blue for higher than 4 Å, cyan for 5 Å, green for 6 Å, yellow for 7 Å, orange for 8 Å, red for 9 Å and grey for lower than 9 Å.

4.2. Structural Studies of MSV geminated virus particles

4.2.1. MSV Virus purification and preliminary EM sample analysis

An 8 g batch of leaves successfully yielded enough virions for microscopy experiments after several optimization methods were tried to identify ways of best removing plant cell contaminants as well as retaining the best concentration of virions within the solution and at the same time best avoiding aggregation. Exclusion of maize stems, the particular buffer conditions and centrifugation speed all played an important role in optimizing the purification. The method is described in Chapter 2 (Figure 4.9).

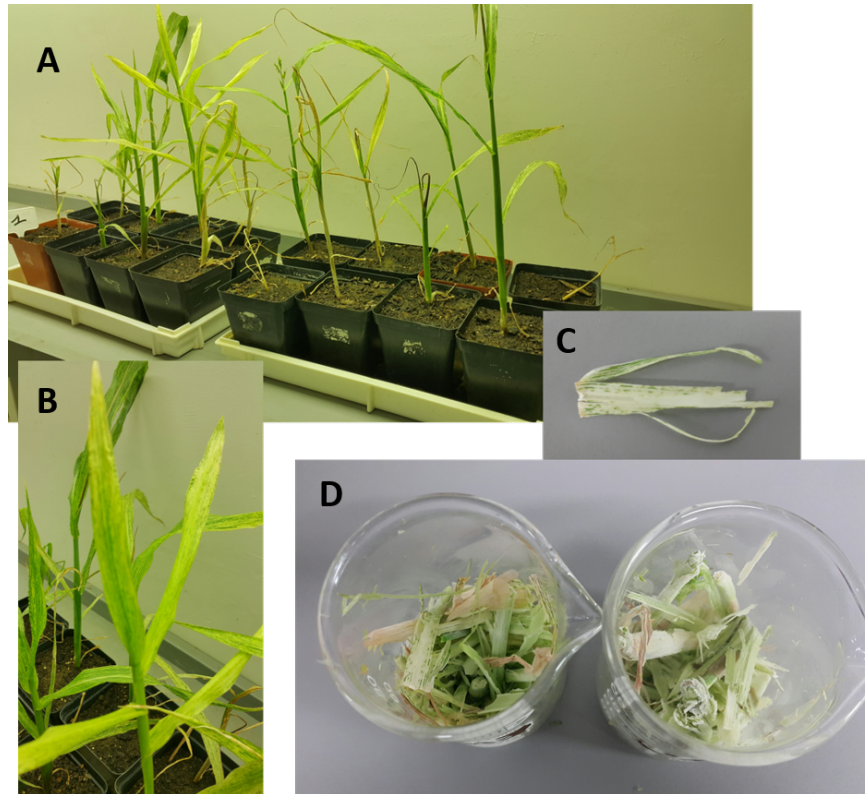


Figure 4.9 The post-infection maize plants, with streaky leaves and unhealthy growth (A and B). (C) is a piece of infected leaf to show the lack of chlorophyll. (D) is an 8g portion of leaves.

To ensure that the virus preparation protocol had worked and that the samples were not damaged by the grid-making procedures, the grids were screened first for each batch. The virus particles were found to be intact with just a very few distorted particles indicating damage. Assured that the sample preparation techniques were consistently functioning as expected, both negative stain and cryo grids were made to examine purified virion samples (Figure 4.10). The MSV virions showed a homogeneous distribution of particles consisting of predominantly bipartite (geminate) assemblies as expected, while occasionally monopartite, tripartite and higher assemblies were also observed. Bipartite assemblies were the main focus of this research and further studied

for their structure, although the other forms of assemblies would be an interesting focus for further research.

Negatively stained MSV virus samples not only provided a primary indication of the sample quality, but also demonstrated the natural appearance of the particles. The negatively stained MSV particles were characterized by the expected twinned-isometric morphology of geminiviruses (Figure 4.10). The particles mainly adopted the flat orientation with the long axis “lying down” on the grids rather than with the short axis “standing up” on the grids, and this is especially obvious in the case of negatively stained particles. The geminate capsid structure has an external diameter of 220 and a length of 380 Å.

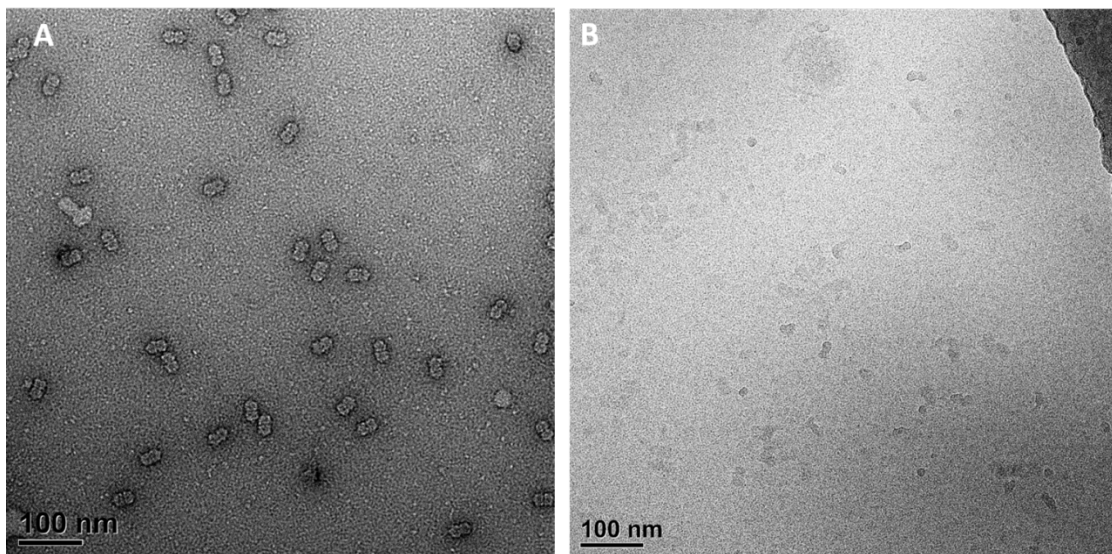


Figure 4.10 MSV geminated virus particle sample checked with negative stained and cryo-EM.

4.2.2. Cryo-EM structure of MSV-gemini via single particle reconstruction

The quality of the preparation was verified in negatively-stained and cryo EM, the particles were intact and level of background debris was low. The sample was at a rather low concentration (could only see less than 5 particles per image on cryo EM), but since the volume of the sample was small, it was not judged worth taking the risk to concentrate the particles, so the sample was applied directly to grids which were plunge- frozen. After plunge-freezing, the virus particles were imaged with cryo-EM, on a Polara microscope (FEI), 2050 movies were collected. A detailed imaging procedure is described in the Materials and Methods chapter. After Motioncorr and CTF correction, particles were picked manually. A total of 5074 picked particles were then classified into 20 2D classes (Figure 4.11). Out of the 20 classes, 17 of them (all except classes 7, 18 and 19) were then classified in 3D into 4 classes with Relion (see Materials and Methods for details). D5 symmetry was applied into the reconstruction, to yield a result in which the particle contained a five-fold rotation symmetry across the long axis of the particle, and a two-fold axis perpendicular to and intersecting this at the particle centre. The 3D classification result is shown in Figure 4.12. Relion also provided an estimate of the angular accuracy of orientations for each 3D class-average. Visual inspection of the classes was also used to determine which 3D classes looked the best and were to be further refined to give the final structure. Figure 4.12 shows the 2D central slice through each of the 3D classes obtained for the MSV geminated particles. Class 1 alone contained 2600 particles, and is the best among all classes judged from Chimera visualisation. The structure was further refined with Relion refinement and post_process, to produce a final D5 structure of geminated MSV virus particles. 2600 particles contributed to the final structure and the final resolution was estimated by the usual gold standard refinement procedure in Relion to be 8.3 Å. This

particle structure is shown in Figure 4.13A and B which show the reconstruction from both a front and a side view.

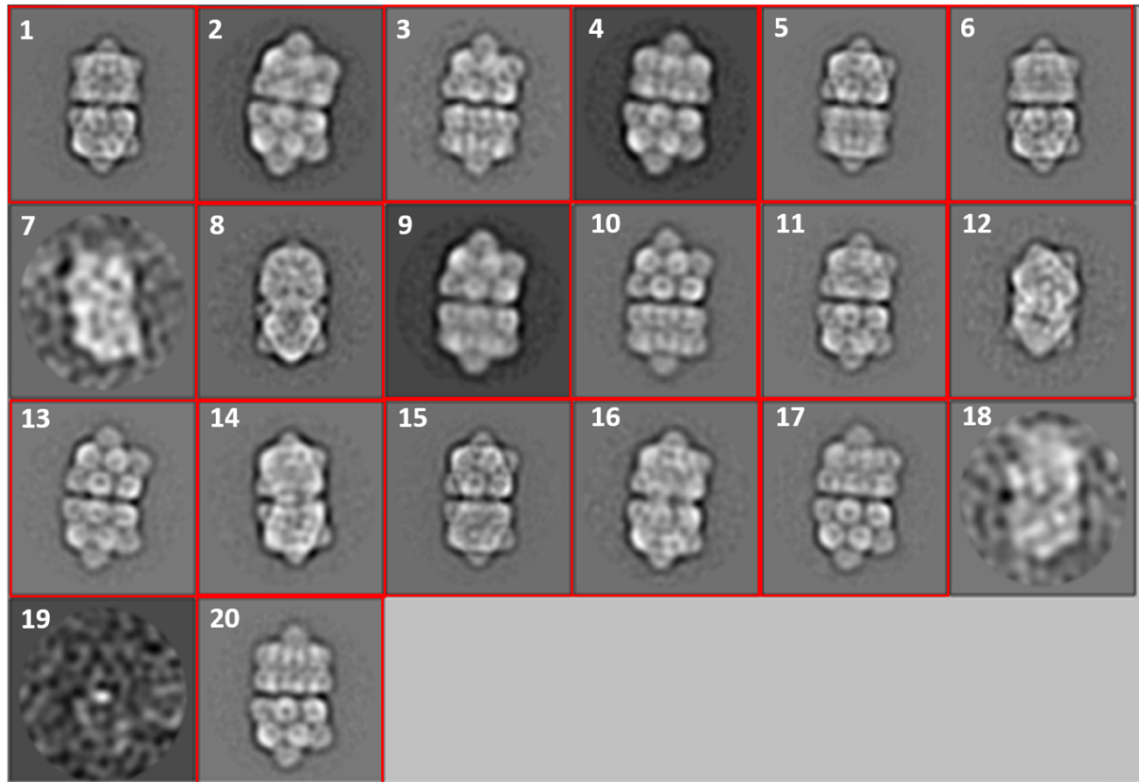


Figure 4.11 Central slices through the 2D classes of the MSV particles. .Except for classes of 7, 18 and 19, all the rest classes (in red boxes) were included into 3D classification.

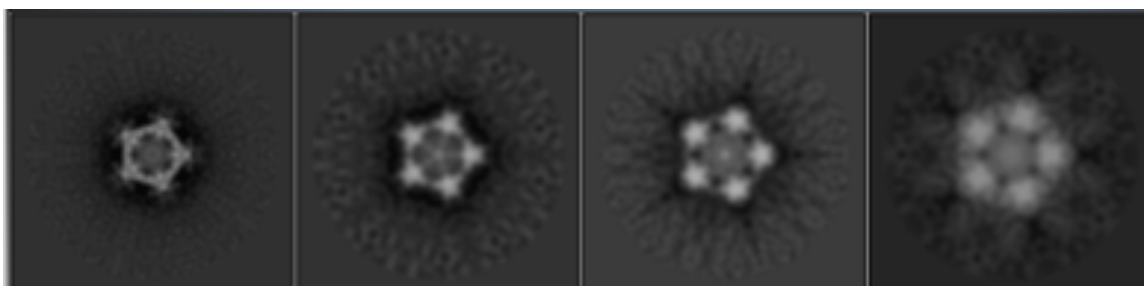


Figure 4.12 2D projections of the four 3D classes of the MSV particles. The first class was taken to be proceeded to refinement.

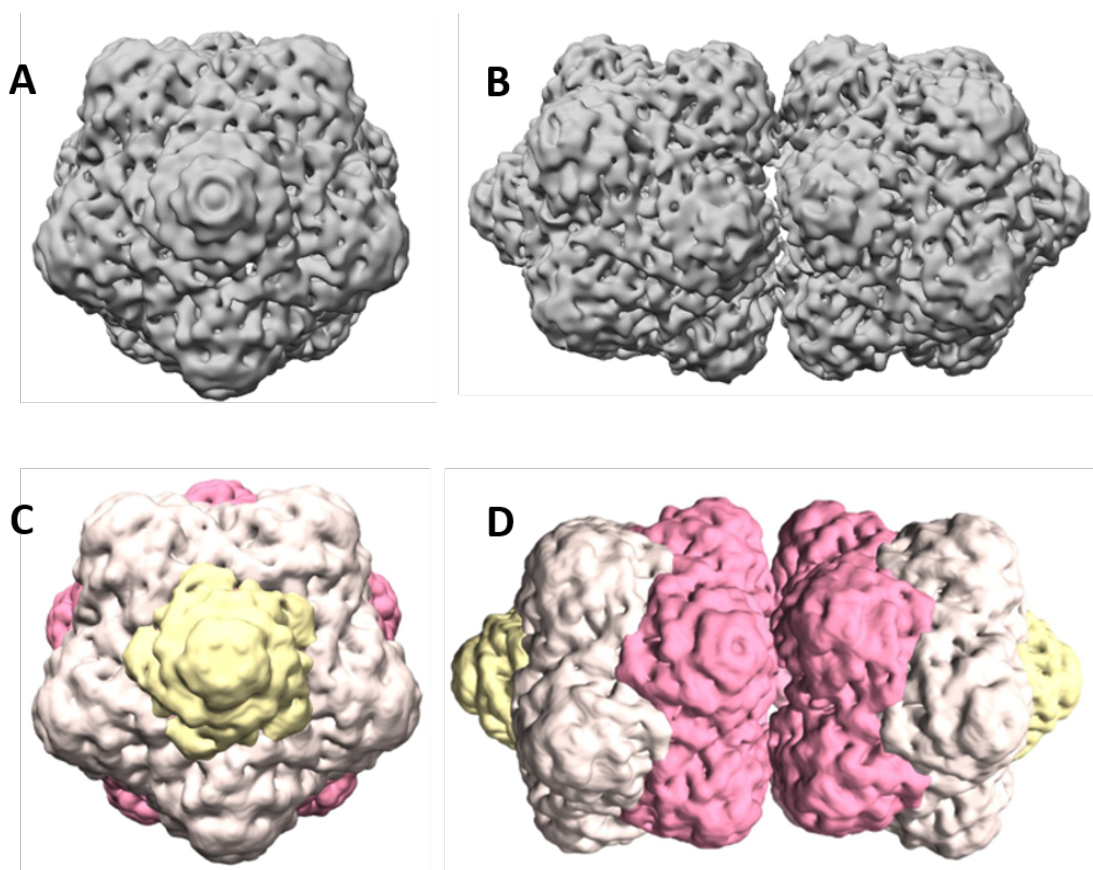


Figure 4.13 The final resolution of the MSV treated as single particles was determined to 8.3 Å. This particle structure is shown in both front view from the 5-fold axis (A) and a side view (B). Three different types of capsomers are coloured in yellow for apical capsomers, white for peripentonal capsomers and pink for equatorial capsomers.

4.2.3. Connection between the two halves of the MSV particles

It is clear from the inspection of the (Figure 4.13) that the capsid has the imposed 52 points group symmetry and as observed by Zhang et al. (2001), each head contains 11 pentameric capsomers that can be further classified into three types of capsomer: apical (at each end of the capsid, shown in yellow), peripentonal (each contacts the apical capsomer, shown in white) and equatorial (connecting the two heads of the particle, shown in pink) (Figure 4.13). Each capsomer is assumed to contain five capsid

proteins (CPs). The apical capsomers at the two ends of the particle follow the strict 5-fold symmetry which coincides with the long axis of the particle while the peripentonal capsomers do not appear to possess exact 5-fold symmetry, but are presumably distorted by the force from the asymmetrical local environment. An oval, rather than a round vertex as that of the apical capsomers' gives the first clear indication of the asymmetry (Figure 4.13). The equatorial capsomers also show an obvious breakdown in local symmetry. From the three-dimensional image reconstruction (Figure 4.14) we can clearly see a gap between the two heads of the MSV particle with specific connections made with the capsid proteins forming some bridges that link the heads. A closer inspection of the equatorial region of the geminate capsid shows that the connections between the two heads consist of five areas of tubular density, with each one linking the two opposed equatorial capsomers from the two halves (Figure 4.14). These protein bridges, together with the genome encapsidated inside the particles ensure a well-defined structure and the robust construction of the virions. The geminated MSV particle structure thus provides a novel system for studying the mechanism by which CPs and encapsidated genomes assemble to form a unique virus particle.

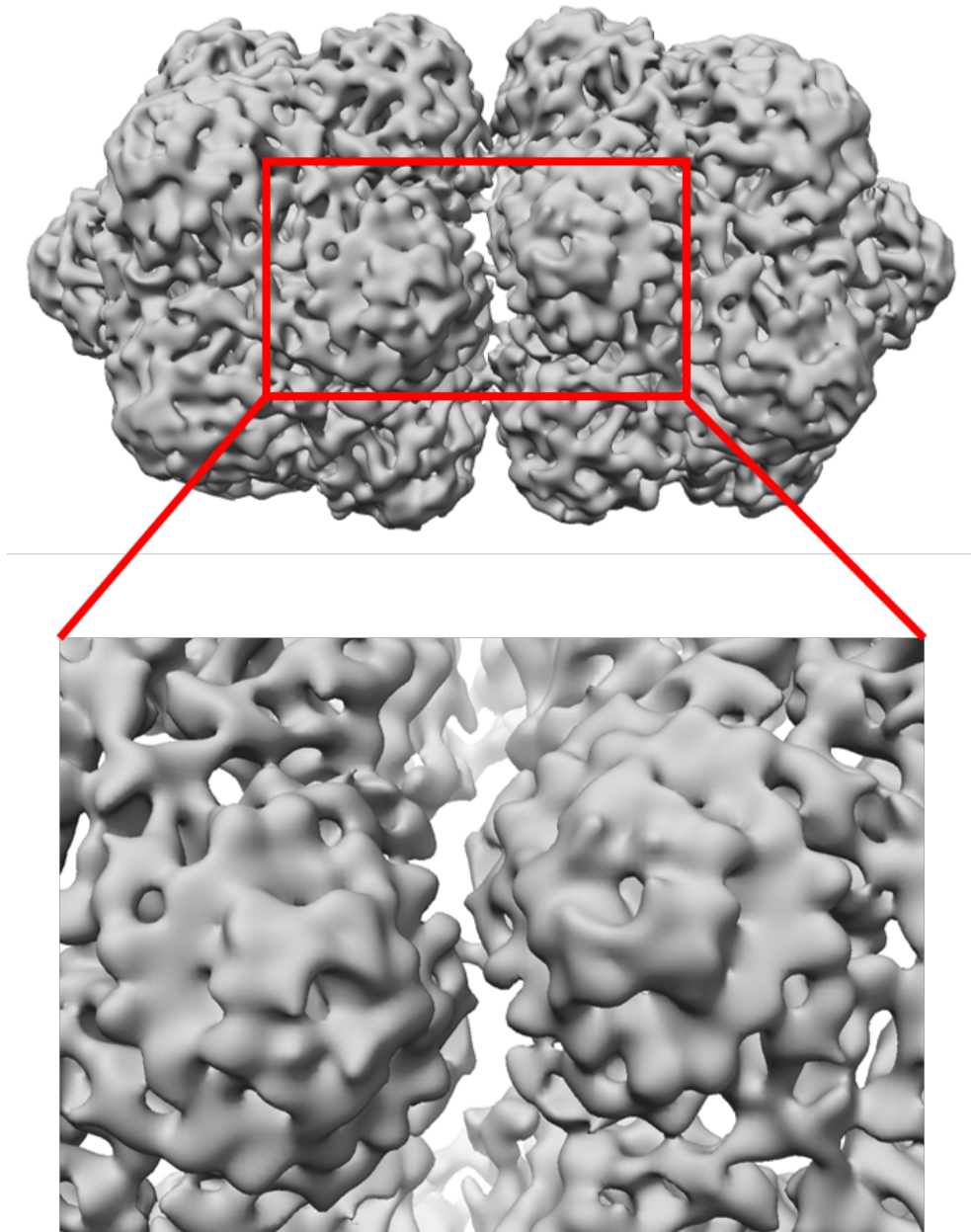


Figure 4.14 A gap between the two heads of the MSV particle is shown with density of bridges linking the halves of the virus.

4.2.4. Localized reconstruction of MSV

The resolution obtained when refining the particles with D5 symmetry was markedly lower than expected for a well behaved rigid particle possessing exact D5 symmetry (based on recent experience within the laboratory). To try to understand if this reflected, for instance, intrinsic flexibility between the two halves of the particles, a localized reconstruction was carried out (Ilca et al. 2015). During this procedure, local areas, which in this case were the two heads from every single particle, were extracted and treated as single particles to further classified and reconstructed, thereby doubling the number of particles in the analysis. For each of these particles the procedure assigns a starting position and orientation based on the overall values for the geminate particle. These particles can then be classified and, if the data are sufficiently robust, their orientations and positions can be refined. The sub-particles were further refined with Relion refinement and post_process, and therefore the final structure of localized reconstructed half-geminated MSV virus particles was produced. The final resolution of MSV structure was determined to 7.4 Å by this method (Figure 4.15). The localized reconstruction of the half particles was duplicated and either half was fitted into the 8.3 Å EM map to generate a whole particle. The localize-reconstructed structure shows significantly more detailed features than that of structure reconstructed only as single particles.

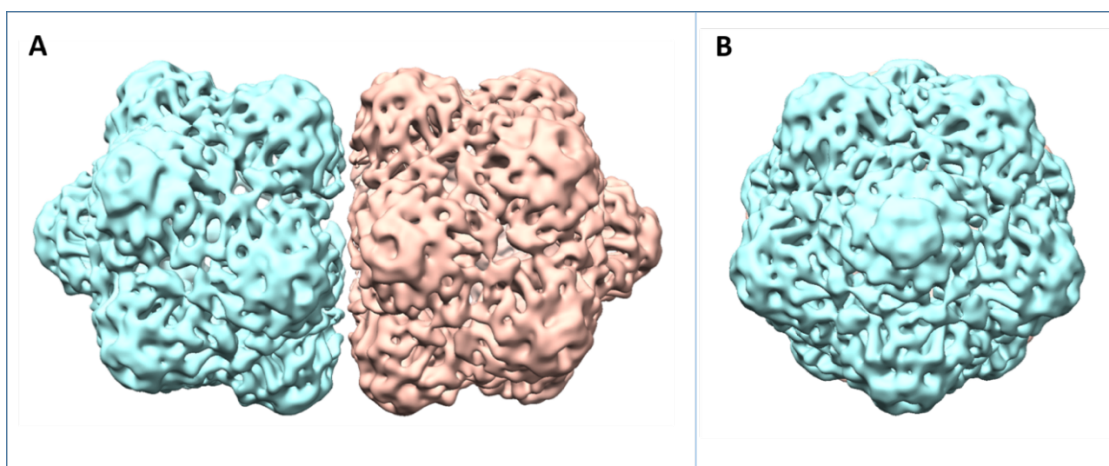


Figure 4.15 Localize reconstructed structure of MSV determined to resolution of 7.4 Å, shown in side view (A) and top view from five-fold axis (B).

4.2.5. Rotation offset of the two halves of MSV particles

The two halves of the geminated MSV virus particles offset from each other by a certain angle. This offset may help maximizes interactions between components of the two halves. Thus, it would be interesting to know by what degree one half of the particle is offset from the other half. The measurement is done by first aligning the EM map generated from the localized reconstruction (described in the latter paragraph) to the left side of the full particle generated from single particle reconstruction, and followed by a flip of the half model. The half model was rotated to fit for the left half of the geminated particle, and by this process the offset degree was measured to be 12°, as this was the angle it took to fit in the map by rotating (Figure 4.16). As mentioned above, the offset of two halves may help maximizes interactions between them, and through this specific degree, the 5 bridges between the adjacent capsomers were formed.

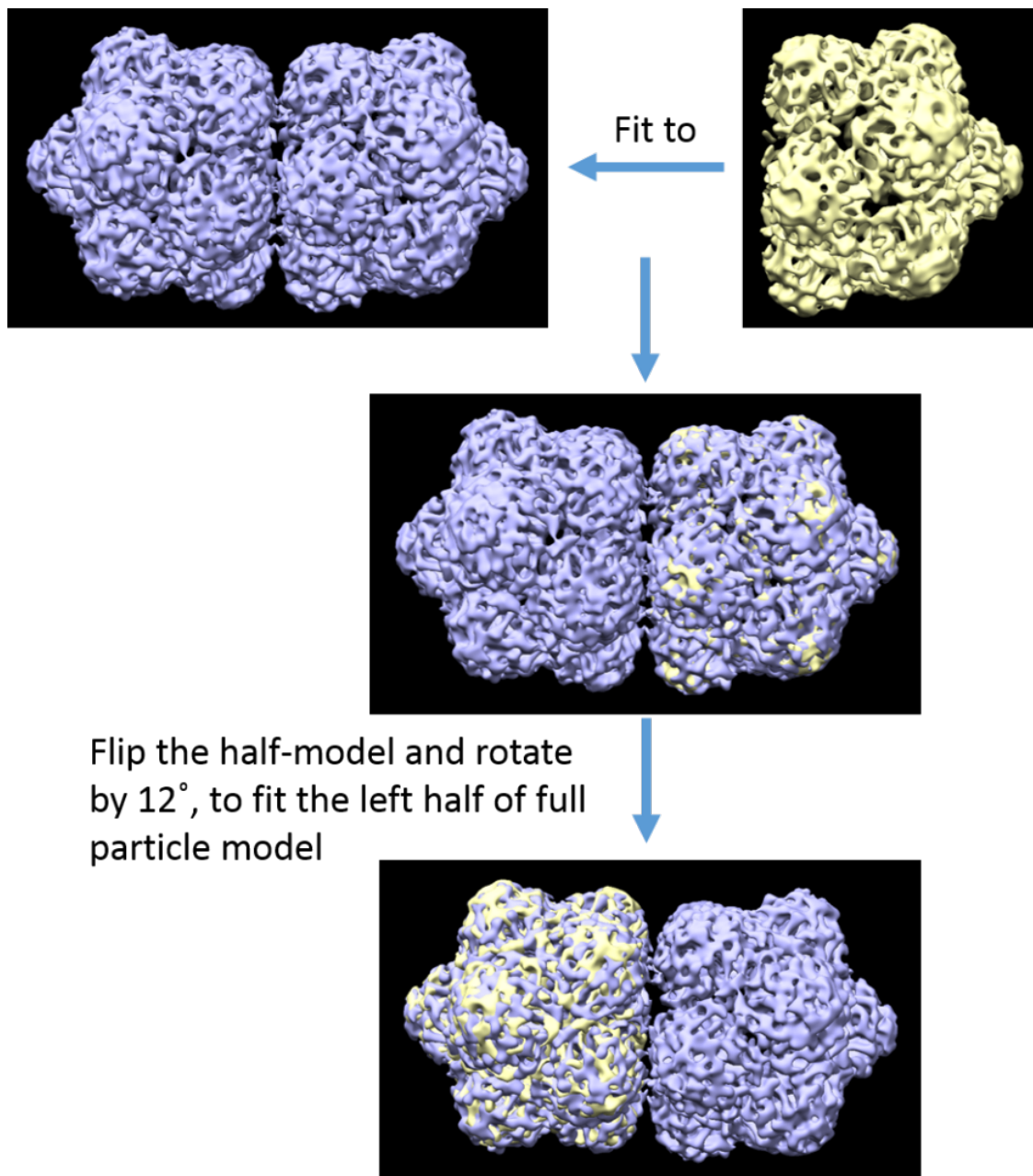


Figure 4.16 Method to illustrate by what degree one half of the particle is offset from the other half. The half head structure generated from localized reconstruction (coloured in yellow) was aligned with the right side of the full particle generated from single particle reconstruction (coloured in purple). After flipping and rotating for 12° , the half model can be fitted with the left half of the full virus model thus to measure out the offset degree.

4.2.6. Resmap analysis of MSV localize reconstructed EM structure

ResMap (Kucukelbir et al. 2014) was used to visualise the local resolution of the EM 3D density maps (Figure 4.17). Comparing with the MSV_VLPs, the overall distribution of the resolution on MSV geminated particles follows a different principle. The apical capsomers possess the highest resolution, with more of colour blue for resolution higher than 4 Å and cyan for 5 Å, while the equatorial capsomers show the most of grey areas representing for resolution lower than 9 Å, and the resolution for peripentonal capsomers in between. This may due to the flexibility of the connection between the two heads, which may apply force and twist the equatorial capsomers by certain extent, and thus introduce difference into individual particles.

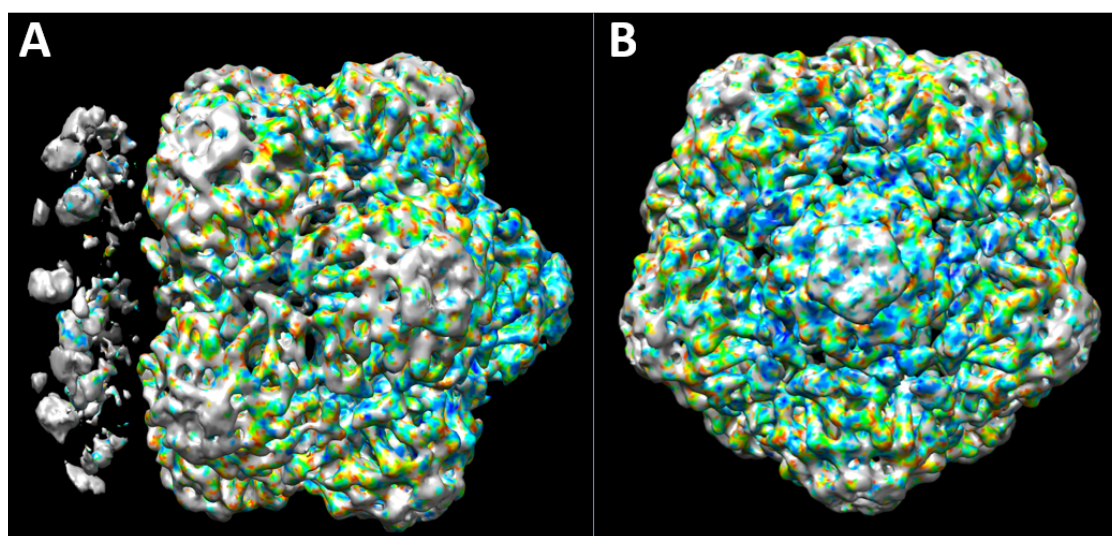


Figure 4.17 ResMap analysis of localize reconstructed structure of half the geminated particle of MSV virus. The colour-code is based on the local resolution of the structure. The colours that represents for resolution ranges are: Blue for higher than 4 Å, cyan for 5 Å, green for 6 Å, yellow for 7 Å, orange for 8 Å, red for 9 Å and grey for lower than 9 Å.

4.2.7. Fitting STNV CP into localize reconstructed MSV EM density

For one of each type of capsomers, 5 STNV CPs were manually fitted in COOT to generate a pentamer. During fitting, the β -sheets provided important points of reference for placing the molecules. For the apical capsomer, 5 monomers were placed in symmetrical pentagonal positions, while in the peripentonal and equatorial capsomers, the 5 monomers were placed in an asymmetric way, especially the equatorial one (Figure 4.18). After the 3 types of pentamers were generated, they were further fitted into the MSV density using Chimera (Figure 4.19). During fitting, 10 solutions were asked for and the best fitting solution (the one with the highest cc value) was adopted. The size difference between the two proteins results in density that has no model fitted which decreases the cc value. For the best fitting solutions for all three types of pentamers (Figure 4.19), the correlation coefficient (CC) value of each were listed in Table 4. .

Table 4.1 Correlation coefficient (CC) value of the three types of pentamers generated from STNV CP monomers fitted onto MSV virus localize reconstructed structure using Chimera.

Type of Pentamers	CC value
Apical	0.75
Peripentonal	0.76
Equatorial	0.74

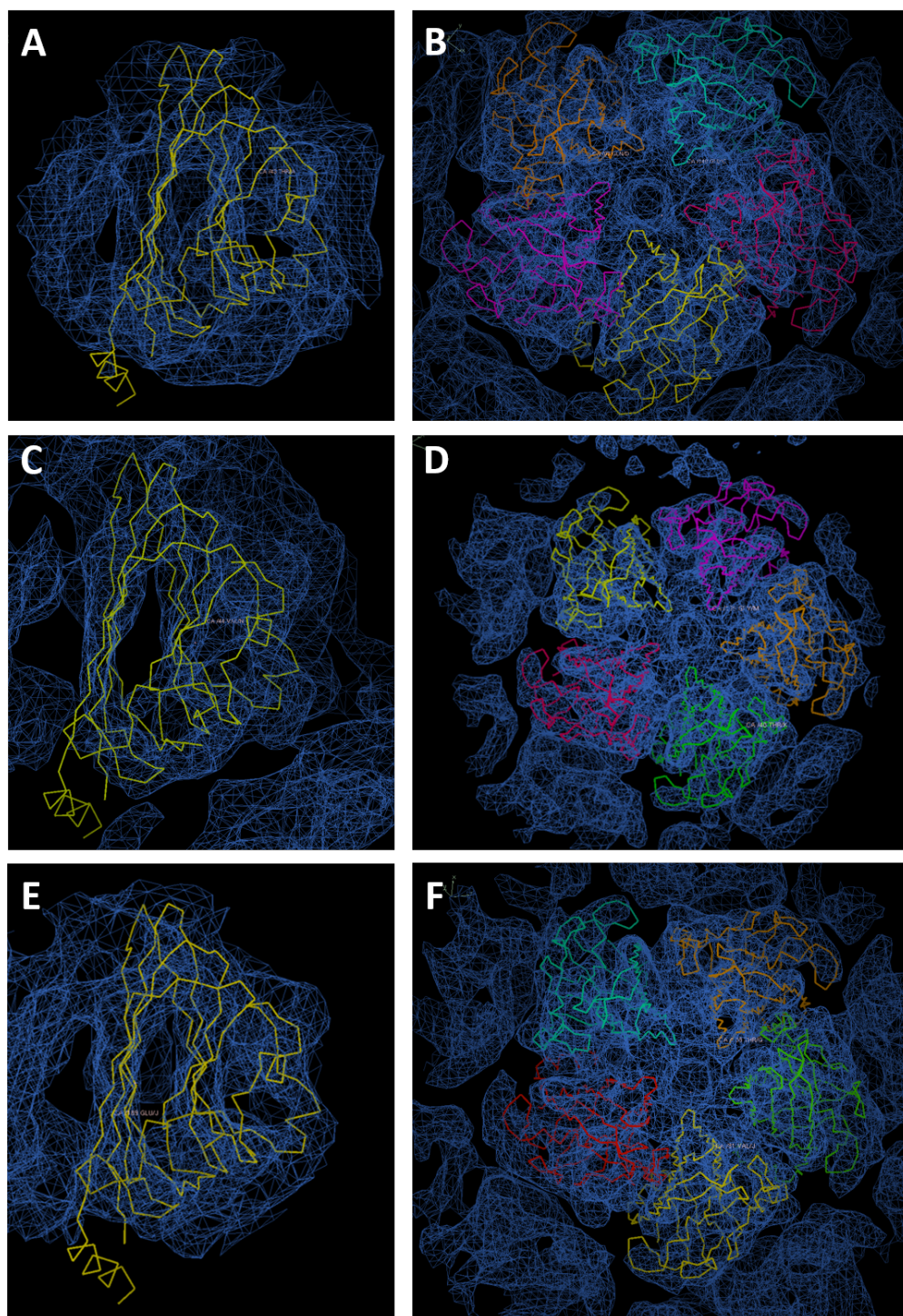


Figure 4.18 The fitting of the atomic monomer model from STNV (ribbon) into the EM density (mesh) of MSV virus to generate three types of pentamers shown from both side view and top view of three different capsomers. A and B for apical capsomers, C and D for peripentonal capsomers and E and F for equatorial capsomers.

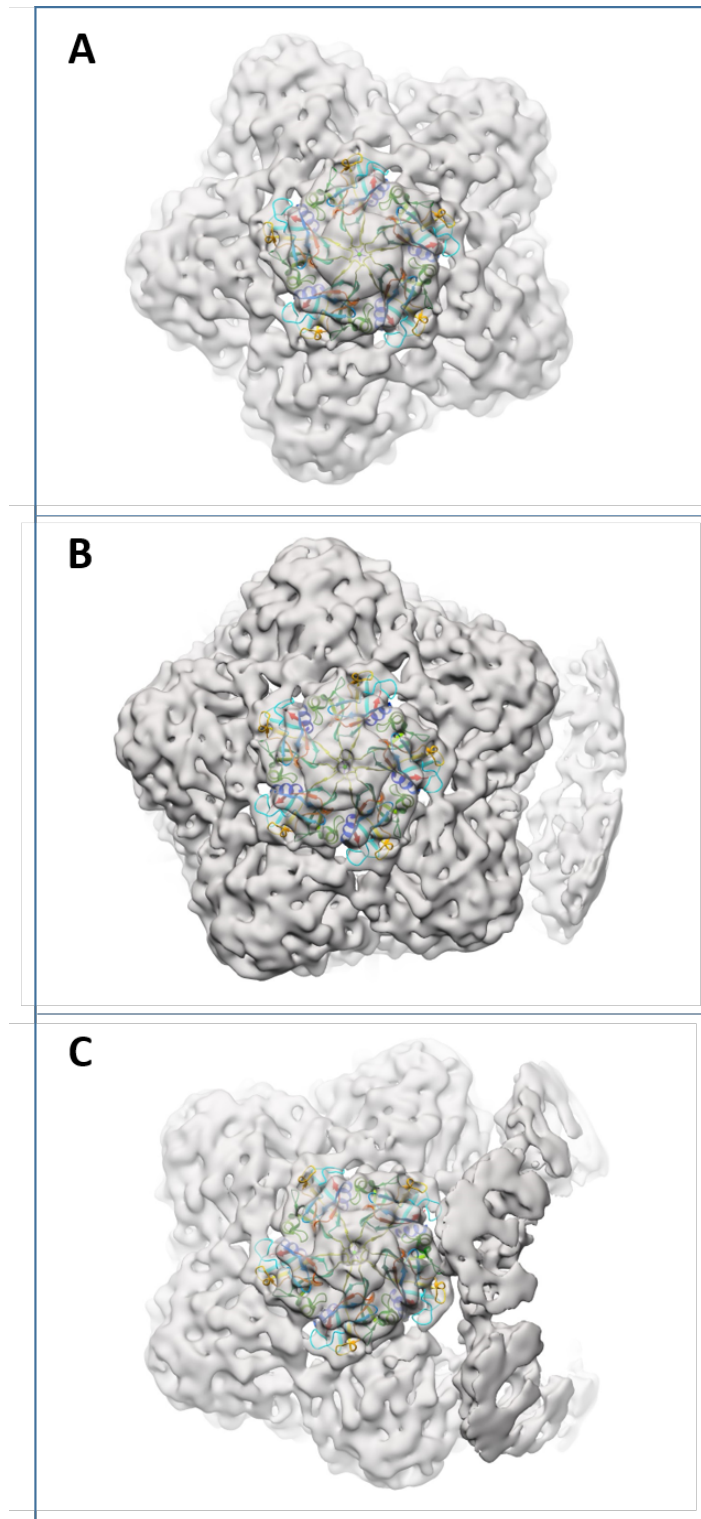


Figure 4.19 The fitting of the atomic pentamer models from STNV into the EM density of MSV virus shown from both side view of three different capsomers. A for apical capsomer, B for peripentonal capsomer and C for equatorial capsomer.

4.3. Discussion

Cryo-EM is becoming an increasingly powerful tool for structure determination of large biological molecules, especially for viruses where they offer icosahedral or other symmetries. The progress in using direct electron detectors has also pushed the resolution boundaries for structure determination using cryo-EM. Higher symmetry, larger particle numbers, homogeneity of the sample can all help to improve the resolution. Both of the two structures of MSV-VLP and MSV virus in a geminated state have both suffered from firstly, a lack of particle numbers that have contributed to the final structure and secondly, homogeneity of the sample, especially the MSV-VLP. For the MSV-VLP sample, the heterogeneous nature of the particles reduced the percentage of the particles that can be used to generate the structure since the majority of the particles picked from the sample were excluded during the 2D and 3D classification. This may be due to the randomness of the nucleic acid the MSV-CP bound to and enclosed inside the particles. One possible way to improve the homogeneity of the sample is by designing and expressing in large scale a RNA of a certain length while the MSV-CP was expressed with *E. coli*. For the geminated MSV virus particles, the particle number was also limited, which may be insufficient for reconstruction of a structure of atomic level. However, different from the case of the VLPs, the main problem with the sample was the concentration of particles frozen on grids was very low. For each of the image taken, only an average of 2 particles can be found. This was not easy to solve from a later stage if the sample volume was low unless the preparation for virus sample were made differently. Another factor limiting the resolution was the flexibility on the joint area of the two halves in the middle of the geminated MSV. This was on a very large extent eliminated with the localized reconstruction by taking either half of the virus as a single sub-particle.

The *Geminiviridae* is a special family of virus which have elongated, geminate capsids with two incomplete T=1 icosahedra joined at the missing vertex. Except for some case of beomoviruses, the genome of the geminiviruses are single components between 2500–3100 nucleotides. The appearance of morphologically distinct assemblies of MSV suggests it may be a solution of packing genome. From the two EM structures, difference is shown when the contour level was both at 0.0087, at which level the density for genome of the virus was shown (Figure 4.20). However, for the MSV VLP, no density from the inside particle can be seen no matter how the contour level was set. One possible suggestion for the assembly of the geminated feature of the virus is to be able to enclose the genome, which may be too big for a single T=1 icosahedral capsid to enclose. When no genome was present while MSV-CP was expressed, the protein formed into a thermodynamically stable structure of icosahedral particles. However, when the CP was packed with genome, a comparably thermodynamically unstable structure of geminated structure was preferred. An experiment of adding supplementary RNA or ssDNA into the capsid protein expression system to test for the assembly may serve more direct evidence.

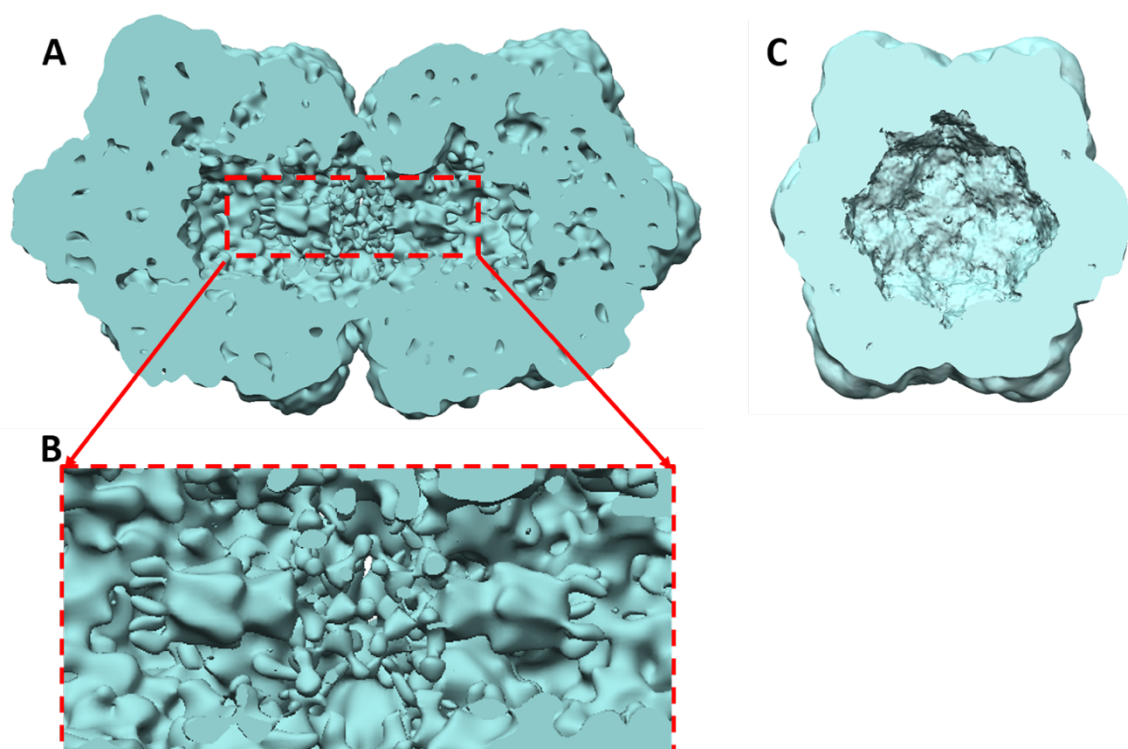


Figure 4.20 Inside density of MSV geminated particle (A) and VLP (B) shown at a contour level of 0.0087. Density for genome can be seen from the virus particles, with a closer view in C, while no density inside VLPs can be seen.

Chapter 5.

Conclusions and Future Work

5.1. Overview

Although described as “organisms at the edge of life” (Rybicki 1990) due to the lack of key cell characteristics, viruses are more abundant than all life forms put together (Delwart 2007), can infect all types of life forms, from animals and plants to microorganisms, including bacteria and archaea (Koonin et al. 2006) and are found in almost every ecosystem on earth (Delwart 2007). Information on virus structures has been growing rapidly, aided first by the development of the electron microscope, then by X-ray crystallography, and now again mainly by the latest generation of advanced electron microscopes. In parallel computational methods and with the growth of molecular biology and virology have advanced massively so that we now have a substantial database of virus structures at high resolution (see for instance Rossmann 2013). Around half of all virus families have an icosahedral arrangement, for example satellite tobacco necrosis virus (STNV), poliovirus, bacteriophage PhiX174 and parvoviruses (Mateu 2013), also the viruses studied in this thesis, bluetongue virus (BTV) and, in part, maize streak virus (MSV) VLP. The MSV geminated virus particles are characterized as a twinned pseudo-icosahedral capsid (W. Zhang et al., 2001). Since the first atomic viral structure was published for tomato bushy stunt virus at 2.9 Å resolution in 1978 (Harrison et al. 1978), nearly 8000 viral related protein structures have been determined (Bassetto et al. 2016). The majority of these were determined by X-ray crystallography, however, the increasing power of cryo-EM as a structural biology tool, marks this out as a rapidly advancing technique, especially due to the improvement in detector technology and computational capabilities (Kühlbrandt 2014). In this thesis, both of these methods were used to solve structures from different viruses. BTV VP5 structures at high and low pHs were solved by X-ray crystallography, both to atomic resolution, which provided insights into the membrane penetration procedure of non-enveloped virus from a structural level. The MSV VLP (single icosahedral particles) and the geminated MSV virion structures were solved

to resolutions of 9.5 Å and 7.4 Å, respectively, by cryo-EM. These structures helped the understanding of the assembly of geminiviruses. More detailed conclusions and future work for each of the two projects are described in the following two paragraphs.

5.2. Conclusions and Future work on BTV VP5 structures

Bluetongue (BT) disease in ruminates is caused by Bluetongue virus (BTV) and has a major impact, both agricultural and economic. BTV encodes 7 structural proteins (VP1-VP7) that form a complex structure of a subcore (VP3), core (VP3+VP7) (Grimes et al., 1998), and outer capsid (VP2 and VP5) layers. Cell entry utilizes clathrin-mediated endocytosis and pH-dependent penetration which involves the dissociation of the outer capsid followed by the release of the core in to the cytosol (Forzan, Marsh, & Roy, 2007). Understanding the mechanism by which the virus enters into host cells may help develop strategies for disease control.

Serotypes of BTV were aligned and phylogenetic groups defined. Tests for expression of VP5 from different phylogenetic groups of BTVs were carried out and the bluetongue virus serotype 15 (BTV15) showed good yields and solubility and was carried on to crystallization. Crystals grow in both high pH (pH9) and a low pH (pH 6) condition. X-ray crystallography was used to determine VP5 structures from both of these crystal forms and revealed, in atomic detail, a novel mechanism of pH dependant conformation. The VP5 crystal at high pH belong to the P1 space group, packed with 9 monomers in a unit cell. Three trimers can be visualized in the crystal unit cell (Figure 3.19A). This structure contains an unfurling domain sitting on top of an anchoring domain, and it shows a conserved surface for the subunit interfaces in the trimer, but a more variable outer surface. In contrast, the low pH VP5 structure maintained all the of anchoring domain, but the majority of unfurling domain after $\alpha 5$ helix was unresolved due to disorder or disassociation. The remaining N-terminal unfurling domain shows a transformation from

a β -sheet to an α -helix conformation and more significantly a change and re-arrangement of 3 bent helices into one long straight helix protruding out of the molecule. The low pH anchoring domain turns the both the last C-terminal β -sheet and the middle loop between $\alpha 15$ and $\beta 8$ into helices. Two histidine residues (H386, H387) within the beta-meander motif displayed changes in the side chain conformation between the two different pHs. Superimposition of the structures at different pHs suggests a H386 conformation change may result in movement of the unfurling domain helix by pushing the conserved E316 residue in the beam helix $\alpha 12$ which contacts the anchoring domain. The pH dependent conformational changes are backed up by the recently reported cryo-EM structure of BTBV-1 (Zhang et al. 2016) and our low pH structure agrees well with that observed in the low pH cryo-EM map. Our high resolution low pH structure clarifies that the protruding objects observed in the low pH cryo EM structure are the stem helices $\alpha 3-5$, whereas they were previously wrongly assumed to be dagger domain helices. Our low pH crystal structure fits well into the cryo-EM map, as shown in Figure 3.33 confirming that our high resolution crystal structure is biologically relevant, and that the distinguishing features we observe are likely to be genuine.

One mystery that still remains in our low pH structure, is that the majority of the unfurling domain is unresolved in the structure. Whether it is completely disordered but still packed in the volume of the crystal not occupied by ordered protein, or whether the “invisible” part was cleaved prior to crystallisation, needs further experiments to clarify. Recrystallisation followed by silver stained gel analysis of dissolved crystals should answer this question. To further investigate the process of conformational change and membrane penetration future work should probably switch to electron microscopy and tomography, with possible experiments including tomographic analysis of particles with liposomes or lipid nanodisks at different pHs.

5.3. Conclusion and Future Work of the MSV CP structures

Geminiviruses are plant viruses infecting a variety of crops, including cassava, cotton, legumes, maize, tomato and okra, leading to great economic losses and posing a serious threat to food security during the past two decades, particularly in the tropics and in subtropical areas (Krupovic et al. 2009). They encompass seven genera, one of them is *Mastrevirus*. Viruses from this genus are found primarily in the Old World, transmitted by leafhoppers and infect both monocotyledonous and dicotyledonous plants. Maize streak virus (MSV) is the type member of *Mastrevirus* that infects maize and causes the devastating maize streak disease (MSD) (Harkins et al., 2009).

The capsid protein (CP) is the only structural protein of MSV and forms a geminated viral particle composed of two incomplete T=1 icosahedra of dimensions 220 x 380 Å with fivefold symmetry axis coincident with the long axis of the particle, so that the CPs are arranged with quasi-icosahedral symmetry (Zhang et al., 2001). When the CP is expressed in an *E. coli* expression system, T=1 icosahedral VLPs were found instead of the geminated ones. I therefore determined structures of these two types of particles to provide evidence that might clarify their assembly. The VLPs from *E. coli*, form icosahedral particles with size of 200 ± 10 Å. A cryo-EM structure of the MSV VLP was determined to 9.5 Å resolution, with 1453 particles contributing to the final structure. Intact MSV viruses were purified from MSV-infected maize leaves from a collaborating lab in South Africa. Single particle reconstruction methods lead to a structure at 8.3 Å resolution, with 2600 particles contributing to the final structure. Structural features such as the connections between the two halves of the MSV particles can be seen from the EM density and a 12° offset of two halves was determined. To avoid the intrinsic flexibility between the two halves of the particles,

a localized reconstruction improved the resolution to 7.4 Å by treating each half of the geminated particle (one incomplete icosahedral half) as a new single particle. STNV CP, as the most similar structure that could be found in the PDB was used to fit into the density for a pentamer of MSV VLP and the three types of pentamers found in the geminated MSV viruses. Real space correlations coefficients were all above 73% indicating a biologically relevant, yet not perfect, fit.

The currently published structure for the MSV virus was at 25 Å, thus our structures represent a major improvement and reveal many more details of the CP. More could be done in the future to improve the resolution. For the MSV VLP sample, the major problem is likely to be the heterogeneous nature of the particles which resulted in only being able to find a small number of similar particles to include in the reconstruction. Introducing RNA or single stranded DNA (ssDNA) of a certain length may help to fix the size of the VLPs. Other methods such as searching for a better buffer condition or a purification procedure which can separate the particles by size more accurately might also be useful. For the MSV virus sample purified from maize plants, the main problem with the sample was the concentration of particles frozen on grids was very low. Each movie taken contained only two particles on average. A purification method which yields a more concentrated sample should be explored and different types of grids can be tried (for instance coated in a single layer of graphene). Multiple blotting on a single grid may also improve the concentration. Simply collecting more data may also help with enlarging the data size thus increase the resolution. Further data processing techniques should also be explored to get the very best result out of the existing data.

References

- Aktories, K. et al., 2010. Current Topics in Microbiology and Immunology *Volume 347 Series Editors Current Topics in Microbiology and Immunology*,
- Aricescu, A.R. & Assenberg, R., 2006. research papers Eukaryotic expression : developments for structural proteomics research papers. , pp.1114–1124.
- Bansal, O.B. et al., 1998. Membrane organization of bluetongue virus nonstructural glycoprotein NS3. *Journal of Virology*, 72(4), pp.3362–3369.
- Bassetto, M. et al., 2016. Structural biology in antiviral drug discovery. *Current Opinion in Pharmacology*, 30, pp.116–130.
- Battye, T.G.G. et al., 2011. research papers iMOSFLM : a new graphical interface for diffraction- image processing with MOSFLM research papers. , pp.271–281.
- Beaton, A.R. et al., 2002. The membrane trafficking protein calpactin forms a complex with bluetongue virus protein NS3 and mediates virus release. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20), pp.13154–13159.
- Bernal, J.D. & Fankuchen, I., 1941. X-Ray and Crystallographic Studies of Plant Virus Preparations. Iii. *The Journal of general physiology*, 25, pp.147–165.
- Berrow, N.S. et al., 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications 35(6).
- Bhattacharya, B. & Roy, P., 2008. Bluetongue virus outer capsid protein VP5 interacts with membrane lipid rafts via a SNARE domain. *Journal of virology*, 82(21), pp.10600–12.
- Binshtein, E. & Ohi, M.D., 2015. Cryo-Electron Microscopy and the Amazing Race to Atomic Resolution. *Biochemistry*, 54, pp.3133–3141.
- Bock, K.R., Guthrie, E.J. & Woods, R.D., 1974. Purification of maize streak virus and its relationship to viruses associated with streak diseases of sugar cane and *Panicum maximum*. *Annals of Applied Biology*, 77(3), pp.289–296.

- Bos, L., 1999. Beijerinck's work on tobacco mosaic virus: historical context and legacy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 354(1383), pp.675–85.
- Bosque-perez, N.A., 2000. Eight decades of maize streak virus research. *Virus res*, 71, pp.107–121.
- Boulton, M.I. et al., 1993. Replication of maize streak virus mutants in maize protoplasts: evidence for a movement protein. *Virology*, 192(1), pp.85–93.
- Boulton, M. et al., 1989. Mutational analysis of the virus-sense genes of maize streak virus. *J. Gen. Virol.*, 70(1989), pp.2309–2323.
- Boulton, M.I. et al., 1989. Specificity of Agrobacterium-mediated delivery of maize streak virus D N A to members of the Gramineae pp.31–40.
- Boyce, M., Celma, C.C.P. & Roy, P., 2012. Bluetongue virus non-structural protein 1 is a positive regulator of viral protein synthesis. *Virology Journal*, 9(1), p.178. Available at: Virology Journal.
- Breitbart, M. & Rohwer, F., 2005. Here a virus, there a virus, everywhere the same virus? *Trends in Microbiology*, 13(6), pp.278–284.
- Brookes, S.M., Hyatt, A.D. & Eaton, B.T., 1993. Characterization of virus inclusion bodies in bluetongue virus-infected cells. *Journal of General Virology*, 74(3), pp.525–530.
- Canchaya, C. et al., 2003. Phage as agents of lateral gene transfer. *Current Opinion in Microbiology*, 6(4), pp.417–424.
- Carpenter, S., Mellor, P.S. & Torr, S.J., 2008. Control techniques for Culicoides biting midges and their application in the U.K. and northwestern Palaearctic. *Medical and Veterinary Entomology*, 22(3), pp.175–187.
- Carvalho, C.M. et al., 2008. NSP-interacting GTPase. *Plant signaling & behavior*, 3(9), pp.752–754.
- Chandran, K., Farsetta, D.L. & Nibert, M.L., 2002. Strategy for nonenveloped virus entry: a hydrophobic conformer of the reovirus membrane penetration protein micro 1 mediates membrane disruption. *Journal of virology*, 76(19), pp.9920–9933.
- Chen, S. et al., 2013. High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy*, 135, pp.24–35.
- Chowda-Reddy, R. V. et al., 2009. Characterization of the cassava geminivirus transcription activation protein putative nuclear localization signal. *Virus Research*, 145(2), pp.270–278.

- Crick, F.H, Watson, J.D., 1956. Structure of Small Viruses. *Nature*, 177, pp.473–475.
- Crowfoot, D.S., 1945. X-Ray Crystallographic Measurements on a Single Crystal of a Tobacco Necrosis Virus Derivative. *Nature*, 155, pp.504–505.
- Day, P.M., Lowy, D.R. & Schiller, J.T., 2003. Papillomaviruses infect cells via a clathrin-dependent pathway. *Virology*, 307(1), pp.1–11.
- De Clercq, K. et al., 2009. Emergence of bluetongue serotypes in Europe, Part 2: The occurrence of a BTV-11 strain in Belgium. *Transboundary and Emerging Diseases*, 56(9-10), pp.355–361.
- Delwart, E.L., 2007. Viral metagenomics. *Reviews in Medical Virology*, 17(2), pp.115–131.
- Denisova, E. et al., 1999. Rotavirus capsid protein VP5* permeabilizes membranes. *Journal of virology*, 73(4), pp.3147–3153.
- Di Emidio, B. et al., 2004. Efficacy and safety studies on an inactivated vaccine against bluetongue virus serotype 2. *Vet Ital*, 40(4), pp.640–644.
- Diprose, J.M. et al., 2002. Translocation portals for the substrates and products of a viral transcription complex: The bluetongue virus core. *EMBO Journal*, 20(24), pp.7229–7239.
- Dormitzer, P.R., Grandi, G. & Rappuoli, R., 2012. Structural vaccinology starts to deliver. *Nature Publishing Group*, 10(12), pp.807–813.
- Eaton, B., Hyatt, A. & Brookes, S., 1990. The replication of bluetongue virus. *Curr Top Microbiol Immunol*, 162, pp.89–118.
- Eberhard-karls, Mainou, B. & Dermody, T.S., 2016. HHS Public Access. pp.91–119.
- Ebert, D.H. et al., 2002. Cathepsin L and cathepsin B mediate reovirus disassembly in murine fibroblast cells. *Journal of Biological Chemistry*, 277(27), pp.24609–24617.
- Emsley, P. & Lohkamp, B., 2010. research papers Features and development of Coot research papers. , pp.486–501.
- Erdmann, J.B. et al., 2010. Replicative intermediates of maize streak virus found during leaf development. *Journal of General Virology*, 91(4), pp.1077–1081.
- Evans, G. & Djinovic, K., 2003. research papers SAD phasing with triiodide, softer X-rays and some help from radiation damage research papers. (2001), pp.1429–1434.
- Evans, P., 2006. Scaling and assessment of data quality. *Acta Crystallographica Section D*, D62, pp.72–82.

- Fallis, A. et al., 2013. Tomato bushy stunt virus at 2.9 Å resolution. *Journal of Chemical Information and Modeling*, 53(9), pp.1689–1699.
- Feenstra, F. et al., 2014. Bluetongue virus without NS3/NS3a expression is not virulent and protects against virulent bluetongue virus challenge. *Journal of General Virology*, 95(PART 9), pp.2019–2029.
- Fillmore, G.C., Lin, H. & Li, J.K.K., 2002. Localization of the single-stranded RNA-binding domains of bluetongue virus nonstructural protein NS2. *Journal of virology*, 76(2), pp.499–506.
- Fondong, V.N., 2013. Geminivirus protein structure and function. *Molecular Plant Pathology*, 14(6), pp.635–649.
- Forzan, M., Marsh, M. & Roy, P., 2007. Bluetongue virus entry into cells. *J Virol*, 81(9), pp.4819–4827.
- Forzan, M., Wirblich, C. & Roy, P., 2004. A capsid protein of nonenveloped Bluetongue virus exhibits membrane fusion activity. *Proc Natl Acad Sci U S A*, 101(7), pp.2100–2105.
- Fricks, C.E. & Hogle, J.M., 1990. Cell-induced conformational change in poliovirus: externalization of the amino terminus of VP1 is responsible for liposome binding. *Journal of virology*, 64(5), pp.1934–1945.
- Gibbs E, Greiner E, Taylor W, Barber T, et al, 1983. Isolation of bluetongue virus serotype 2 from cattle in Florida: Serotype of bluetongue virus hitherto unrecognized in the Western Hemisphere. *Am J Vet Res*, (44), p.2226.
- Gibbs EPJ, Tabachnick WJ, Holt TJ, S. DE, 2008. U.S. Concerns over Bluetongue. *Science*, 320, pp.872–874.
- Ginalski, K., Elofsson, A. & Fischer, D., 2003. 3D-Jury: a simple approach to improve protein structure predictions. , 19(8), pp.1015–1018.
- Goodman, R.M., 1977. Single-stranded DNA genome in a whitefly- transmitted plant virus. *Virology*, 83(1), pp.171–9.
- Gouet, P. et al., 1999. The Highly Ordered Double-Stranded RNA Genome of Bluetongue Virus Revealed by Crystallography. *Cell*, 97, pp.481–490.
- Grimes, J.M. et al., 1998. The atomic structure of the bluetongue virus core. *Nature*, 395(6701), pp.470–478.
- Grimsley, N. et al., 1989. DNA transfer from Agrobacterium to Zea mays or Brassica by agroinfection is dependent on bacterial virulence functions. *Mol Gen Genet*, 217, pp.309–316.

- Guirakhoo, F., Catalan, J. & Monath, T., 1995. Adaptation of bluetongue virus in mosquito cells results in overexpression of NS3 proteins and release of virus particles. *Arch Virol.*, 140(5), pp.967–74.
- Gutierrez, C., 2000. DNA replication and cell cycle in plants: learning from geminiviruses. *The EMBO journal*, 19(5), pp.792–799.
- Hallenberger, S. et al., 1990. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature*, 26(360), pp.358–61.
- Hardy WT, Price DA: Sore muzzle of sheep. *J Am Vet Med Assoc* 1952, 120:23–25.
- Harkins, G.W. et al., 2009. Dating the origins of the maize-adapted strain of maize streak virus, MSV-A. *Journal of General Virology*, 90(12), pp.3066–3074.
- Harrison, B.D. et al., 1977. Plant viruses with circular single-stranded DNA. *Nature*, 270, pp.760–762.
- Harrison, S.C. et al., 1978. Tomato bushy stunt virus at 2.9A resolution. *Nature*, 276, p.368.
- Hartitz, M.D., Sunter, G. & Bisaro, D.M., 1999. The tomato golden mosaic virus transactivator (TrAP) is a single-stranded DNA and zinc-binding phosphoprotein with an acidic activation domain. *Virology*, 263(1), pp.1–14.
- Hassan, S.H. et al., 2001. Expression and functional characterization of bluetongue virus VP5 protein: role in cellular permeabilization. *Journal of virology*, 75(18), pp.8356–8367.
- Hassan, S.S. & Roy, P., 1999. Expression and functional characterization of bluetongue virus VP2 protein: role in cell entry. *Journal of virology*, 73(12), pp.9832–42.
- Hewat, E. et al., 1992. 3-D reconstruction of bluetongue virus tubules using cryoelectron microscopy. *Journal of Structural Biology*, 108(1), pp.35–48.
- Hildebrand, A. et al., 2009. PREDICTION REPORT Fast and accurate automatic structure prediction with HHpred ". pp.128–132.
- Hofmann, M.A. et al., 2008. Genetic characterization of toggenburg orbivirus, a new bluetongue virus, from goats, Switzerland. *Emerging Infectious Diseases*, 14(12), pp.1855–1861.
- Hong, Y. et al., 1996. Resistance to geminivirus infection by virus-induced expression of dianthin in transgenic plants. *Virology*, 220(1), pp.119–27.
- Hormuzdi, S.G. & Bisaro, D.M., 1995. Genetic analysis of beet curly top virus: Examination of the roles of L2 and L3 genes in viral pathogenesis. *Virology*, 206(2), pp.1044–1054.

- Horscroft, N.J. & Roy, P., 2000. NTP binding and phosphohydrolase activity associated with purified bluetongue virus non-structural protein NS2. *Journal of General Virology*, 81(8), pp.1961–1965.
- Huismans, H., van Dijk, A.A. & Els, H.J., 1987. Uncoating of parental bluetongue virus to core and subcore particles in infected L cells. *Virology*, 157(1), pp.180–188.
- Huxley, H.E. & Zubay, G., 1960. The structure of the protein shell of turnip yellow mosaic virus. *Journal of Molecular Biology*, 2(4), pp.189–IN5.
- Hyatt, A., Eaton, B. & Brookes, S., 1989. The release of bluetongue virus from infected cells and their superinfection by progeny virus. *Virology*, 173(1), pp.21–34.
- Ilca, S.L. et al., 2015. cryomicroscopy images of macromolecular complexes. *Nature Communications*, 6, pp.1–8.
- Jahn, R. & Sudhof, T., 1999. M Embrane F Usion and E Xocytosis. *Annual Reviews in Biochemistry*, pp.777–810.
- Jané-valbuena, J. et al., 2002. Sites and Determinants of Early Cleavages in the Proteolytic Processing Pathway of Reovirus Surface Protein, 76(10), pp.5184–5197.
- Jenckel, M. et al., 2015. Complete coding genome sequence of putative novel bluetongue virus serotype 27. *Genome announcements*, 3(2), pp.16–17.
- Johnson, J.E., 2010. *Cell entry by non-enveloped viruses*,
- Jupin, I. et al., 1994. Movement of tomato yellow leaf curl geminivirus (TYLCV): involvement of the protein encoded by ORF C4. *Virology*, 204(1), pp.82–90.
- Kabsch, W., 2010. XDS. *Biological Crystallography*, D66, pp.125–132.
- Kabsch, W., 2010. Integration , scaling , space-group assignment and post-refinement research papers. , pp.133–144.
- Kabsch, W., 1988. Automatic indexing of rotation diffraction patterns. *J. Appl. Cryst.*, 21, pp.67–71.
- Kalia, M. & Jameel, S., 2011. Virus entry paradigms. *Amino Acids*, 41, pp.1147–1157.
- Kämper, N. et al., 2006. A membrane-destabilizing peptide in capsid protein L2 is required for egress of papillomavirus genomes from endosomes. *Journal of virology*, 80(2), pp.759–68.
- Kantardjieff, K.A., 2003. Matthews coefficient probabilities : Improved estimates for unit cell contents of proteins , DNA , and protein – nucleic acid complex crystals. , pp.1865–1871.

- Kar, A.K., Bhattacharya, B. & Roy, P., 2007. Bluetongue virus RNA binding protein NS2 is a modulator of viral replication and assembly. *BMC molecular biology*, 8, p.4.
- Kar, A.K. & Roy, P., 2003. Defining the Structure-Function Relationships of Bluetongue Virus Helicase Protein VP6 Defining the Structure-Function Relationships of Bluetongue Virus Helicase Protein VP6. 77(21), pp.11347–11356.
- Karlson, P. & Luscher, M., 1959. Structure of Poliomyelitis Virus. *Nature*, 183(4653), pp.55–56.
- Kay, L., 1986. W. M. Stanley's crystallization of the tobacco mosaic virus, 1930-1940. *Isis*, 77(288), pp.450–72.
- Kendrew, J.C. et al., 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181(4610), pp.662–666.
- Kim, K.-S. & Lee, K.-W., 1992. Geminivirus-Induced Macrotubules and Their Suggested Role in Cell-to-Cell Movement. *Cytology and Histology*, 82, pp.664–669.
- Koonin, E. V. & Starokadomskyy, P., 2016. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, 59, pp.125–134.
- Koonin, E. V, Senkevich, T.G. & Dolja, V. V, 2006. The ancient Virus World and evolution of cells. *Biology direct*, 1(29), p.27.
- Krupovic, M., Ravantti, J.J. & Bamford, D.H., 2009. Geminiviruses : a tale of a plasmid becoming a virus. , 11, pp.1–11.
- Kucukelbir, A., Sigworth, F. & Tagare, H., 2014. The Local Resolution of Cryo-EM Density Maps. *Nat Methods*, 11(1), pp.63–65.
- Kühlbrandt, W., 2014. The Resolution Revolution. , 343(March), pp.2014–2016.
- Lai, J. et al., 2009. RKP, a RING finger E3 ligase induced by BSCTV C4 protein, affects geminivirus infection by regulation of the plant cell cycle. *Plant Journal*, 57(5), pp.905–917.
- Lane, S.W. et al., 2011. Construction and crystal structure of recombinant STNV capsids. *Journal of Molecular Biology*, 413(1), pp.41–50.
- Laufs, J. et al., 1995. Geminivirus replication: Genetic and biochemical characterization of rep protein function, a review. *Biochimie*, 77(10), pp.765–773.
- Lazarowitz, S. & Beachy, R., 1999. Viral movement proteins as probes for intracellular and intercellular trafficking in plants. *The Plant cell*, 11(4), pp.535–548.

- Leslie, A.G.W. & Powell, H.R., 2007. Processing diffraction data with mosflm. In *Evolving Methods for Macromolecular Crystallography*. pp. 41–51.
- Lett, J.-M. et al., 2002. Spatial and temporal distribution of geminiviruses in leafhoppers of the genus *cicadulina* monitored by conventional and quantitative polymerase chain reaction. *Phytopathology*, 92(1), pp.65–74.
- Li, X. et al., 2013. Electron counting and beam-induced motion correction enable near atomic resolution single particle cryoEM. *HHS Public Access*, 10(6), pp.584–590.
- Liebert, M.A. et al., 2001. Rate Matrices for Analyzing Large Families of Protein Sequences. 8(4), pp.381–399.
- Liu, H., Boulton, M.I. & Davies, J.W., 1997. Maize streak virus coat protein binds single- and double-stranded DNA in vitro. *Journal of General Virology*, 78, pp.1265–1270.
- Liu, H., Oparka, K.J. & Davies, J.W., 2001. Interaction of the movement and coat proteins of. *The Journal of general virology*, 82(Pt 1), pp.35–44.
- Lourenco, S. & Roy, P., 2011. Supplement: In vitro reconstitution of Bluetongue virus infectious cores. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33), pp.13746–51.
- Lucy, A.P. et al., 1996. Tissue specificity of Zea mays infection by maize streak virus. *Molecular Plant-Microbe Interactions*, 9(1), pp.22–31.
- Maclachlan, N.J. et al., 2009. The Pathology and Pathogenesis of Bluetongue. *Journal of Comparative Pathology*, 141(1), pp.1–16.
- Maclachlan, N.J., 2010. Global Implications of the Recent Emergence of Bluetongue Virus in Europe. *Vet Clin Food Anim*, 26, pp.163–171.
- Maclachlan, N.J. & Mayo, C.E., 2013. Potential strategies for control of bluetongue, a globally emerging, Culicoides-transmitted viral disease of ruminant livestock and wildlife. *Antiviral Research*, 99(2), pp.79–90.
- Maclachlan, N.J. & Mayo, C.E., 2013. Potential strategies for control of bluetongue, a globally emerging, Culicoides-transmitted viral disease of ruminant livestock and wildlife. *Antiviral Research*, 99(2), pp.79–90.
- Marsh, M. & Helenius, A., 2006. Virus entry: Open sesame. *Cell*, 124(4), pp.729–740.
- Martin, D.P., Willment, J. a & Rybicki, E.P., 1999. Evaluation of Maize Streak Virus Pathogenicity in Differentially Resistant Zea mays Genotypes. *Phytopathology*, 89(8), pp.695–700.
- Martinez-costas, J. et al., 1997. Bluetongue Virus VP6 Protein Binds ATP and Exhibits an RNA-Dependent ATPase Function and a Helicase Activity That Catalyze the Unwinding of Double-Stranded RNA Substrates. 71(10), pp.7220–7226.

- Mateu, M.G., 2013. Assembly , stability and dynamics of virus capsids. *Archives of Biochemistry and Biophysics*, 531(1-2), pp.65–79.
- Matsuo, E. & Roy, P., 2011. Bluetongue virus VP1 polymerase activity in vitro: Template dependency, dinucleotide priming and cap dependency. *PLoS ONE*, 6(11).
- Matthews, B.W., 1968. Solvent Content of Protein. *J. Mol. Biol*, 33, pp.491–497.
- Mccoy, A.J. et al., 2007. research papers Phaser crystallographic software research papers. pp.658–674.
- McGarry, R.C. et al., 2003. A novel Arabidopsis acetyltransferase interacts with the geminivirus movement protein NSP. *The Plant cell*, 15(7), pp.1605–1618.
- Meiswinkel, R. et al., 2004. The taxonomy of Culicoides vector complexes - unfinished business. *Veterinaria italiana*, 40(3), pp.151–159.
- Mendez, I.I. et al., 2003. Digestion pattern of reovirus outer capsid protein3 determined by mass spectrometry. , 311, pp.289–304.
- Mertens, P., 2004. The dsRNA viruses. *Virus Research*, 101(1), pp.3–13.
- Mesfin, T., Den Hollander, J. & Markham, P.G., 1995. Feeding activities of Cicadulina mbila (Hemiptera: Cicadellidae) on different host-plants. *Bulletin of Entomological Research*, 85(03), p.387.
- Mills-Lujan, K. & Deom, C.M., 2010. Geminivirus C4 protein alters Arabidopsis development. *Protoplasma*, 239(1-4), pp.95–110.
- Mindell, J.A. & Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. 142, pp.334–347.
- Monjane, A.L. et al., 2011. Recombination hotspots and host susceptibility modulate the adaptive value of recombination during maize streak virus evolution. *BMC evolutionary biology*, 11(1), p.350.
- Moyer, C.L. & Nemerow, G.R., 2011. Viral weapons of membrane destruction : variable modes of membrane penetration by non-enveloped viruses. *Current Opinion in Virology*, 1(1), pp.44–49.
- Moyer, C.L. et al., 2011. Functional genetic and biophysical analyses of membrane disruption by human adenovirus. *Journal of virology*, 85(6), pp.2631–41.
- Mubin, M. et al., 2010. The hypersensitive response induced by the V2 protein of a monopartite begomovirus is countered by the C2 protein. *Molecular Plant Pathology*, 11(2), pp.245–254.
- Mulherkar, N. et al., 2012. HHS Public Access. , 477(7364), pp.340–343.

- Mumtsidu, E. et al., 2007. Structural features of the Bluetongue virus NS2 protein. *Journal of Structural Biology*, 160(2), pp.157–167.
- Noad, R. & Roy, P., 2009. Bluetongue vaccines. *Vaccine*, 27(SUPPL. 4), pp.86–89.
- Noueiry, A.O., Lucas, W.J. & Gilbertson, R.L., 1994. Two proteins of a plant DNA virus coordinate nuclear and plasmodesmal transport. *Cell*, 76(5), pp.925–932.
- Owens, R.J., Limn, C. & Roy, P., 2004. Role of an arbovirus nonstructural protein in cellular pathogenesis and virus release. *Journal of virology*, 78(12), pp.6649–6656.
- Padidam, M., Beachy, R.N. & Fauquet, C.M., 1996. The role of AV2 (“precoat”) and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology*, 224(2), pp.390–404.
- Park, J. et al., 2011. The arabidopsis thaliana homeobox gene ATHB12 is involved in symptom development caused by geminivirus infection. *PLoS ONE*, 6(5).
- Parker, J.S. et al., 2001. Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. *Journal of virology*, 75(8), pp.3896–3902.
- Pasumarthy, K.K., Choudhury, N.R. & Mukherjee, S.K., 2010. Tomato leaf curl Kerala virus (ToLCKeV) AC3 protein forms a higher order oligomer and enhances ATPase activity of replication initiator protein (Rep/AC1). *Virology journal*, 7, p.128.
- Perutz, M.F. et al., 1960. Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*, 185(4711), pp.416–422.
- Peterschmitt, M. et al., 1992. Detection of maize streak virus antigens over time in different parts of maize plants of a sensitive and a so-called tolerant cultivar by ELISA. *Annals of Applied Biology*, 121(3), pp.641–653.
- Pickles, R.J. et al., 1998. Limited entry of adenovirus vectors into well-differentiated airway epithelium is responsible for inefficient gene transfer. *Journal of virology*, 72(7), pp.6014–23.
- Piroux, N. et al., 2007. Geminivirus pathogenicity protein C4 interacts with Arabidopsis thaliana shaggy-related protein kinase AtSK??, a component of the brassinosteroid signalling pathway. *Virology*, 362(2), pp.428–440.
- Poranen, M.M., Daugelavičius, R. & Bamford, D.H., 2002. Common Principles in Viral Entry. *Annual Review of Microbiology*, 56(1), pp.521–538.
- Preiss, W. & Jeske, H., 2003. Multitasking in Replication Is Common among Geminiviruses. , 77(5), pp.2972–2980.

- Purcell, E.M., 1956. © 1956 Nature Publishing Group. *Nature*.
- Ramachandran, G.N., Ramakrishnan, C. & Sasisekharan, V., 1963. Stereochemistry of Polypeptide Chain Configurations. *Journal of Molecular Biology*, 7(1), pp.95–99..
- Ratinier, M. et al., 2016. Bluetongue virus NS4 protein is an interferon antagonist and a determinant of virus virulence. *Journal of Virology*, 90(11), pp.JVI.00422–16.
- Ren, J. et al., 2015. Structures of Coxsackievirus A16 Capsids with Native Antigenicity : Implications for Particle Expansion , Receptor Binding , and , 89(20), pp.10500–10511.
- Ridgeway, T. & Nw, L., 2000. T-Coffee : A Novel Method for Fast and Accurate Multiple Sequence Alignment.
- Rigden, J.E. et al., 1993. Mutagenesis of the Virion-Sense Open Reading Frames of Tomato Leaf Curl Geminivirus. *Virology*, 193, pp.1001–1005.
- Robert, X. & Gouet, P., 2014. Deciphering key features in protein structures with the new ENDscript server. , 42(April), pp.320–324.
- Roberts, I.M., 1989. Indian cassava mosaic virus: ultrastructure of infected cells. *Journal of General Virology*, 70(1989), pp.2729–2739.
- Rodríguez-Sánchez, et al, 2008. Orbiviruses in the mediterranean basin: updated epidemiological situation of bluetongue and new methods for the detection of BTV serotype 4. *Transboundary and Emerging Diseases*, 55(5-6), pp.205–214.
- Rojas, M.R. et al., 2001. Functional analysis of proteins involved in movement of the monopartite begomovirus, Tomato yellow leaf curl virus. *Virology*, 291(1), pp.110–125.
- Ros, C., Gerber, M. & Kempf, C., 2006. Conformational changes in the VP1-unique region of native human parvovirus B19 lead to exposure of internal sequences that play a role in virus neutralization and infectivity. *Journal of virology*, 80(24), pp.12017–12024.
- Rose, D.J.W., 1978. Epidemiology of maize streak disease. *Annual Review of Entomology*, 23(133), pp.259–282.
- Rossmann, M.G., 2013. Structure of viruses: a short history. *Quarterly reviews of biophysics*, 46(2), pp.133–80.
- Rossmann, M.G., Rao, Z. & Venigalla, B., 2012. Viruses : Sophisticated Biological Machines. , pp.10–12.
- Rott, R. & Siddell, S., 2017. One hundred years of animal virology. *Journal of General Virology*, (1998), pp.2871–2874.

- Roy, P. et al., 1990. Identification of bluetongue virus VP6 protein as a nucleic acid-binding protein and the localization of VP6 in virus-infected vertebrate cells. *Journal of virology*, 64(1), pp.1–8.
- Roy, P., 2009. Bluetongue virus : dissection of the polymerase complex. , 89(Pt 8), pp.1789–1804.
- Rybicki, E., 1990. The classification of organisms at the edge of life or Problems with virus systematics. *Sout*, 86(July), pp.182–186.
- Saegerman, C., Berkvens, D. & Mellor, P.S., 2008. Bluetongue epidemiology in the European Union. *Emerging Infectious Diseases*, 14(4), pp.539–544.
- Saunders, K., Lucy, A. & Stanley, J., 1991. DNA forms of the geminivirus African cassava mosaic virus consistent with a rolling circle mechanism of replication. *Nucleic Acids Research*, 19(9), pp.2325–2330.
- Sauter, N.K., Grosse-kunstleve, R.W. & Adams, P.D., 2004. research papers Robust indexing for automatic data collection research papers. , pp.399–409.
- Savini, G. et al., 2008. Vaccines against bluetongue in Europe. *Comparative Immunology, Microbiology and Infectious Diseases*, 31(2-3), pp.101–120.
- Scheres, S.H.W., 2012. RELION : Implementation of a Bayesian approach to cryo-EM structure determination. *Journal of Structural Biology*, 180(3), pp.519–530.
- Selth, L. a, Randles, J.W. & Rezaian, M.A., 2004. Host responses to transient expression of individual genes encoded by Tomato leaf curl virus. *Molecular plant-microbe interactions : MPMI*, 17(1), pp.27–33.
- Sharma, P. & Ikegami, M., 2010. Tomato leaf curl Java virus V2 protein is a determinant of virulence, hypersensitive response and suppression of posttranscriptional gene silencing. *Virology*, 396(1), pp.85–93.
- Shepherd, D.N. et al., 2010. Maize streak virus: An old and complex “emerging” pathogen. *Molecular Plant Pathology*, 11(1), pp.1–12.
- Shivaprasad, P. V et al., 2005. Promoters , Transcripts , and Regulatory Proteins of Mungbean Yellow Mosaic Geminivirus Promoters , Transcripts , and Regulatory Proteins of Mungbean Yellow Mosaic Geminivirus. , 79(13), pp.8149–8163.
- Skubak, P., Murshhudov, G.N. & Pannu, N.S., 2004. research papers Direct incorporation of experimental phase information in model refinement research papers. *Biological Crystallography*, D60, pp.2196–2201.
- Sperlova, A. & Zendulkova, D., 2011. Bluetongue: A review. *Veterinarni Medicina*, 56(9), pp.430–452.
- Stanley, J. & Latham, J.R., 1992. A symptom variant of beet curly top geminivirus produced by mutation of open reading frame C4. *Virology*, 190(1), pp.506–509.

- Storey, H.H., 1924. The transmission of a new plant virus disease by insects. *Nature*, 114(2859), p.245.
- Stuart, D.I. et al., 1979. Crystal Structure of Cat Muscle Pyruvate Kinase at a Resolution of 2-6 Å. , pp.109–142.
- Sung, Y.K. & Coutts, R.H.A., 1995. Mutational analysis of potato yellow mosaic geminivirus. *Journal of General Virology*, 76(7), pp.1773–1780.
- Sutton, G. et al., 2007. Bluetongue virus VP4 is an RNA-capping assembly line. *Nature structural & molecular biology*, 14(5), pp.449–51.
- Szelke, M. et al., 1982. Potent new inhibitors of human renin. *Nature*, 299(9), pp.555–557.
- Tang, G. et al., 2007. EMAN2 : An extensible image processing suite for electron microscopy. , 157, pp.38–46.
- Teng, K. et al., 2010. Involvement of C4 protein of beet severe curly top virus (Family Geminiviridae) in virus movement. *PLoS ONE*, 5(6), pp.1–8.
- Thorn, A. & Sheldrick, G.M., 2011. ANODE : anomalous and heavy-atom density calculation. , pp.1285–1287.
- Toussaint J.F., Vandenbussche F., Mast J., De Meester L., Goris N., Van Dessel W., Vanopdenbosche E., Kerkhofs P., De Clercq K., Zientara S., Sailleau C., Czaplicki G., Depoorter G., D.J., 2006. No Title. *Vet. Rec*, (2), p.327.
- Tsai, B., 2007. Penetration of Nonenveloped Viruses into the Cytoplasm. *Annu. Rev. Cell Dev. Biol*, 23, pp.23–43
- Vagin, A., 2010. research papers Molecular replacement with MOLREP research papers. , pp.22–25.
- Varagona, M.J. & Raikhel, N. V., 1994. The basic domain in the bZIP regulatory protein Opaque2 serves two independent functions: DNA binding and nuclear localization. *Plant Journal*, 5(2), pp.207–214.
- Varsani, A. et al., 2014. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. *Archives of Virology*, 159(8), pp.2193–2203.
- Vosdingh, R.A., Trainer, D.O. & Easterday, B., 1968. Experimental Bluetongue Disease in White-Tailed Deer. *Can. J. Comp. Med. Vet. Sci*, 32, pp.382–387.
- Walter, T.S. et al., 2005. A procedure for setting up high-throughput nanolitre crystallization experiments . Crystallization workflow for initial screening , automated storage , imaging and optimization. , pp.651–657.

- Ward, B.M. et al., 1997. The geminivirus BL1 movement protein is associated with endoplasmic reticulum-derived tubules in developing phloem cells. *Journal of virology*, 71(5), pp.3726–3733.
- Wartig, L. et al., 1997. Genetic analysis of the monopartite tomato yellow leaf curl geminivirus: roles of V1, V2, and C2 ORFs in viral pathogenesis. *Virology*, 228(2), pp.132–140.
- Watson, J.D. & Crick, F.H., 1953. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361), pp.964–967.
- Wehrfritz, J. et al., 2007. Reconstitution of bluetongue virus polymerase activity from isolated domains based on a three-dimensional structural model. *Biopolymers*, 86(1), pp.83–94.
- Weichenberger, C.X. & Rupp, B., 2014. research papers Ten years of probabilistic estimates of biocrystal solvent content : new insights via nonparametric kernel density estimate research papers. , pp.1579–1588.
- Wiethoff, C.M. et al., 2005. Adenovirus protein VI mediates membrane disruption following capsid disassembly. *Journal of virology*, 79(4), pp.1992–2000.
- Wilson, A.J. & Mellor, P.S., 2009. Bluetongue in Europe: past, present and future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530), pp.2669–2681.
- Wilson, A. & Mellor, P., 2009. Bluetongue in Europe: Vectors, epidemiology and climate change (Parasitology Research (2008) 103, S69-S77 *Parasitology Research*, 104(2), p.489.
- Winn, M.D. et al., 2011. research papers Overview of the CCP 4 suite and current developments research papers. , 4449, pp.235–242.
- Winter, G., 2010. xia2 : an expert system for macromolecular crystallography data reduction. , 2000, pp.186–190.
- Wirblich, C., Bhattacharya, B. & Roy, P., 2006. Nonstructural protein 3 of bluetongue virus assists virus release by recruiting ESCRT-I protein Tsg101. *Journal of virology*, 80(1), pp.460–73.
- Zbyszek, O. & Wladek, M., 1997. Processing of X-Ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, 276(January 1993), pp.306–315.
- Zhang, S.C., Ghosh, R. & Jeske, H., 2002. Subcellular targeting domains of Abutilon mosaic geminivirus movement protein BC1. *Archives of Virology*, 147(12), pp.2349–2363.
- Zhang, W. et al., 2001. Structure of the Maize Streak Virus Geminale Particle. , 477, pp.471–477.

- Zhang, X. et al., 2016. Atomic model of a nonenveloped virus reveals pH sensors for a coordinated process of cell entry. *Nature structural & molecular biology*, 23(1), pp.74–80.
- Zhao, Y. et al., 1994. Deletion and mutational analyses of bluetongue virus NS2 protein indicate that the amino but not the carboxy terminus of the protein is critical for RNA-protein interactions. *Journal of virology*, 68(4), pp.2179–85.
- Zhao, Y., Chapman, D.A.G. & Jones, I.M., 2003. Improving baculovirus recombination. , 31(2).
- Zientara, S. et al., 2010. Bluetongue vaccination in Europe. *Expert Review of Vaccines*, 9(9), pp.989–991.